

**CEP Discussion Paper No 991**

**August 2010**

**(Revised November 2014)**

**Spillovers in Space:  
Does Geography Matter?**

**Sergey Lychagin, Joris Pinkse, Margaret E. Slade  
and John Van Reenen**

## **Abstract**

Using US firm level panel data we simultaneously assess the contributions to productivity of three potential sources of research and development spillovers: geographic, technological, and product market (“horizontal”). To do so, we construct new measures of geographic proximity based on the distribution of a firm’s inventor locations as well as its headquarters. We find that geographic location is important for productivity, perhaps dominating other spillover mechanisms, and that both intra– and inter–regional (counties) spillovers matter. The geographic location of a firm’s researchers is more important than its headquarters. These benefits may be the reason why local policy–makers compete so hard for the location of local R&D labs and high tech workers.

JEL Classifications: O33, L60, C23

Keywords: R&D spillovers, geographic proximity, technological proximity, horizontal proximity, spatial econometrics

This paper was produced as part of the Centre’s Productivity and Innovation Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

## **Acknowledgements**

We would like to thank the co-editor, Chad Syverson, two anonymous referees, Pierre–Philippe Combes, Erwin Diewert, Henry Overman, and numerous seminar participants for helpful comments. Sergey Lychagin and Joris Pinkse thank the Human Capital Foundation for their support of CAPCP. Joris Pinkse’s research is supported by NSF grant SES-0922127. Margaret Slade thanks the Social Sciences and Humanities Research Council of Canada for support. John Van Reenen would like to thank the Economic and Social Research Council for financial support through the Centre for Economic Performance.

Sergey Lychagin, Department of Economics, Central European University. Joris Pinkse, Center for Auctions, Procurements and Competition Policy, Department of Economics, Pennsylvania State University. Margaret E. Slade, Vancouver School of Economics, The University of British Columbia. John Van Reenen, Director of Centre for Economic Performance, London School of Economics, CEPR and NBER.

Published by  
Centre for Economic Performance  
London School of Economics and Political Science  
Houghton Street  
London WC2A 2AE

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means without the prior permission in writing of the publisher nor be issued to the public or circulated in any form other than that in which it is published.

Requests for permission to reproduce any article or part of the Working Paper should be sent to the editor at the above address.

© S. Lychagin, J. Pinkse, M. E. Slade and J. Van Reenen, resubmitted 2014

## I. INTRODUCTION

Firms engage in research and development (R&D) because they anticipate future benefits in the form of lower costs, new methods of production, and new uses for their products. To the extent that they can capture the fruits of their efforts, private and social returns to R&D should be equal.<sup>1</sup> However, when they cannot appropriate the returns, R&D becomes a public good and underinvestment is expected. A knowledge of the magnitude and direction of the problem (i.e., who benefits from the efforts of whom) is therefore a prerequisite for an appropriate design of public policy.

We address the much studied question of the relationship between R&D spillovers and productivity, where a spillover occurs when firm  $i$  benefits from the R&D activities of a different firm  $j$ . There is substantial evidence that spillovers exist. Furthermore, it is commonly agreed that spillovers are larger when firms are *closer*. However, it is not clear what closeness means or how it can be measured. These questions are the focus of our research. Specifically, although we are most interested in geographic proximity, we look at variants of three measures of closeness: proximity in horizontal (product market), technological (patent technology class), and geographic space.<sup>2</sup>

With our focus on geographic proximity, we ask an additional question — Is distance dead? Much has been written on this issue. For example, in her influential book *The Death of Distance*, [Cairncross \(2001\)](#) states that, in a world where information flows electronically and instantaneously, “new ideas will spread faster, leaping borders, and poor countries will have immediate access to information that was once restricted to the industrial world.” On the other hand, the governments of most industrialized nations subsidize the R&D activities of their domestic firms with the hope that subsidies will help those firms become more productive vis-à-vis their foreign rivals.

Our research incorporates a number of novelties that we highlight. First, we assess all three spillover measures simultaneously, which removes the possibility that one measure (say geographic proximity) appears to be important simply because researchers fail to control adequately for the common influence of a second measure (say technological proximity). In other words, firms that are technologically similar might locate in the same region and benefit from each other’s R&D activities. There has been little work that uses the same data and model to evaluate multiple measures of closeness, both individually and jointly.<sup>3</sup>

Second, we refine the measures of closeness, particularly geographic proximity. To illustrate, many previous researchers have used the (single) location of a firm’s headquarters to calculate geographic distance. It is not clear, however, that technological advances are likely to be communicated by CEOs and managers. We hypothesize that inventors are more apt to exchange ideas and we therefore use the distribution of a firm’s (multiple) inventor locations to calculate

---

<sup>1</sup>For optimality, technology markets must also be competitive. In other words, strategic behavior can also distort decisions (see, e.g., [Spence, 1984](#)).

<sup>2</sup>We do not assess vertical proximity because we do not have adequate data.

<sup>3</sup>As we discuss below, although a few researchers evaluate two measures, most assess only one.

proximity. Moreover, we look at the match between firms' R&D activities across regions (counties) and discount that match with a measure of the distance between regions. To our knowledge, no one in the firm-level R&D productivity literature has exploited firms' within country research locations to create a multidimensional multiregional measure of geographic closeness.

As a third contribution, when we find that distance is not dead, we attempt to determine why. In particular, we assess whether geographic spillovers are more important when firms' technological and/or product market R&D activities are closer or whether the effect that we uncover is purely geographic. Finally, by examining whether the estimated half-life distance of the geographical spillover terms have increased over time, we ask: if distance is not dead, is it perhaps dying?<sup>4</sup>

Our research attempts to answer a number of questions that are relevant for economic policy. First and foremost, what is the nature of spillovers? In particular, which firms will benefit from the R&D activities of a given firm and by how much? In the absence of an answer to this question, it is impossible to formulate appropriate public policy towards R&D. Furthermore, it cannot be answered by considering one source of spillovers in isolation. Second, to what extent is knowledge private or public? In other words, can firms capture the lion's share of the rents that are associated with their R&D activities or are they freely available to rivals? This question must be answered if we are interested in assessing under and over investment and the need to correct the associated externalities. Finally, are knowledge flows confined to narrowly defined product markets, technology classes and/or geographic regions? Put differently, are spillovers local or global? This question must be answered if we are concerned with asymmetric patterns of regional or industrial growth and decline and whether growth paths will converge or diverge.

We use U.S. manufacturing firm level accounting data (sales, employment, capital, R&D, etc.) from U.S. Compustat 1980–2000 matched into the U.S. Patent and Trademark Office (USPTO) data from the NBER data archive. To analyze that data, we employ simple versions of the spatial econometric techniques that were developed in [Pinkse et al. \(2002\)](#); [Pinkse and Slade \(2004\)](#). With those techniques, which allow us to handle multiple measures of closeness in a flexible but parsimonious manner, space is broadly defined to include a number of geographic and nongeographic dimensions.

The paper is organized as follows. Section [II](#) discusses the previous literature, Section [III](#) the model, and Section [IV](#) measuring spillovers. The data is discussed in Section [V](#), econometric methods in Section [VI](#), and results in Section [VII](#). Some concluding comments are offered in Section [VIII](#).

## II. PREVIOUS LITERATURE

[Griliches \(1979\)](#) was perhaps the first to recognize the multi-dimensional nature of the spillover

---

<sup>4</sup>Evidence from patent citations data suggests that distance is becoming less important in the international dimension [Griffith et al. \(2011\)](#).

problem. He cast the problem in a production–function framework and discussed alternative hypotheses about the nature of closeness which, in his view, could be horizontal, technological, or vertical. More recently, researchers have recognized that there is another type of closeness that [Griliches \(1979\)](#) did not emphasize — geographic proximity. We discuss each of these possibilities.

*Geographic or learning from neighbors.* Researchers who are located in the same or proximate regions have opportunities to communicate. For example, employees of different firms that are located in close geographic proximity might meet on the golf course or belong to common civic organizations where they discuss and learn from each other’s research activities. In other words, social networks can facilitate learning. In addition, geographic proximity can enable capitalization of complementarities among firms’ research activities that create synergies. Furthermore, R&D agglomeration effects can lead to a better regional infrastructure and attract more highly skilled workers into the region. Finally, when inventors change jobs, taking their ideas with them, they often find a new job in the same geographic region. Geographic spillovers are particularly important since they affect the potential for regional and national convergence. Indeed, if spillovers are highly localized, the prospects for convergence are poor. Such spillovers have been studied by [Adams and Jaffe \(1996\)](#); [Eaton and Kortum \(1996\)](#); [Keller \(2002\)](#); [Orlando \(2004\)](#).

*Horizontal or learning from product–market rivals.* Firms that produce similar products often benefit from each other’s R&D activities. For example, when a pharmaceutical firm introduces a new drug, in the absence of patent protection, rival companies can easily determine its makeup and offer close substitutes. Horizontal spillovers, which are perhaps the most studied, have been evaluated by inter alia [Bernstein \(1988\)](#), [Adams and Jaffe \(1996\)](#); [Orlando \(2004\)](#); [Ornaghi \(2006\)](#).<sup>5</sup>

*Technological or learning from technology–market rivals.* Firms that perform similar types of research can also learn from each other. For example, the discovery of froth flotation, which facilitated the refining of sulfide ores, was an important advance in the mining industry that benefited firms from many different product markets (e.g., copper, lead, and zinc metals, which are used in wiring, batteries, and galvanizing, respectively). Technology–market spillovers were the focus of the work in [Jaffe \(1986\)](#); [Bloom et al. \(2013\)](#).

Finally, firms that are related through a vertical chain might experience technological synergies. We do not evaluate vertical spillovers due to a lack of firm data on buying and selling. Extending our framework to buyer–seller linkages is an important avenue for future research.

Most early spillover studies used data on industrial sectors to examine intersectoral input/output flows or vertical spillovers based on sectoral input/output tables.<sup>6</sup> More recently, however, researchers have used firm level data that allow them to examine such issues as intra

<sup>5</sup>There is also a related literature that examines the relationship between technological diffusion, decentralization, and closeness to the technological frontier (e.g. [Acemoglu et al., 2007](#)).

<sup>6</sup>See, e.g., [Scherer \(1982\)](#); [Griliches and Lichtenberg \(1984\)](#); [Goto and Suzuki \(1989\)](#).

versus inter product market or geographic spillovers. Many of the more recent studies have found significant spillover effects. However, most concentrate on one proximity measure in isolation.<sup>7</sup>

Our research, which is based on firm data in a multi-proximity measure or multi-dimensional spatial setting, takes a production-function approach to the problem. There are of course other approaches. For example, [Bernstein and Nadiri \(1989\)](#) used a dynamic dual framework based on a cost function and a system of factor demands.<sup>8</sup> Still others have used a consumer-surplus framework (see, e.g., the survey by [Griliches, 1992](#)). One advantage of the static production function framework over more structural models is that it requires fewer behavioral assumptions and is therefore subject to less misspecification bias, at least from that source.

There are a number of issues that we have chosen not to cover. For example, we consider private R&D but neglect public and academic activity (for work on the latter, see e.g., [Adams, 1990](#); [Jaffe, 1989](#); [Acs et al., 1992](#); [Adams and Clemmons, 2008](#)). Furthermore, although we exploit patent data to create our measure of geographic proximity, we do not consider patent citations as direct measures of spillovers (as in e.g., [Jaffe et al., 1993](#); [Thompson, 2006](#); [Belenzon and Schankerman, 2013](#)). Third and most importantly, we do not consider vertical or input/output spillovers because we do not have detailed data on interfirm purchases and sales.

### III. THE MODEL

We use a standard production-function framework similar to that employed by [Griliches \(1979\)](#). In particular, output  $Q$  is produced by conventional inputs capital  $K$ , labor  $L$ , and raw materials,  $M$ . In addition to the conventional inputs, output depends on a stock of knowledge or productivity, which we model as Hicks neutral. Firms' R&D activities augment that stock, which depreciates over time.

Specifically, there are  $n$  firms,  $i = 1, \dots, n$ , observed in  $T$  time periods,  $t = 1, \dots, T$ . The production function is

$$(1) \quad Q_{it} = A_{it}\Omega_t U_{it} H_i(x_{it}),$$

where  $Q$  is output,  $A \times \Omega \times U$  is the state of knowledge or productivity, and  $x$  is a vector of conventional inputs. Taking logs we have

$$(2) \quad q_{it} = a_{it} + h_i(x_{it}) + \omega_t + u_{it},$$

where lower case letters denote logarithms. Productivity consists of two parts: a systematic component  $a_{it}$ , and a random component. Furthermore, the random component is subdivided

<sup>7</sup>Exceptions include [Orlando \(2004\)](#); [Adams and Jaffe \(1996\)](#), who assesses product-market and geographical proximity in a discrete framework (e.g., same product market/different geographic region), and [Bloom et al. \(2013\)](#), who assess product and technology-market proximity.

<sup>8</sup>Their dynamic dual approach allows one to endogenize R&D and capital investment decisions through the use of Euler equations. This is a plus if the specified model is correct but a minus if not.

into an aggregate shock,  $\omega_t$ , and an idiosyncratic mean zero shock  $u_{it}$ .

We model knowledge acquisition using a standard capital-accumulation type equation. Specifically, firm  $i$ 's systematic stock of knowledge is

$$(3) \quad S_{it} = (1 - \delta)S_{it-1} + R_{it-1},$$

where  $\delta$  is the depreciation rate and  $R_{it}$  is firm  $i$ 's investment in knowledge in period  $t$ .

We assume that the systematic component of productivity is a weighted average of the R&D activities of all firms

$$(4) \quad a_{it} = \theta S_{it} + \sum_{j \neq i} w_{ij} S_{jt},$$

where  $w_{ij}$  is a weight that corresponds to some notion of the distance between  $i$  and  $j$ . Our objective is to uncover the weighting matrix  $W = [w_{ij}]$ .

We do this as follows. Assume that we have  $K$  distance (really proximity) measures, where each measure produces an  $n \times n$  matrix,  $D^k = [d_{ij}^k]$ , and define the spillover pool associated with  $D^k$  to be

$$(5) \quad S_{-ikt} = \sum_{j \neq i} d_{ij}^k S_{jt}.$$

In other words, rival knowledge stocks can potentially contribute to a firm's productivity but they are attenuated by distance in some metric.

Based on this definition, the simplest parametric weights correspond to

$$(6) \quad a_{it} = \theta S_{it} + \zeta_k S_{-ikt},$$

or

$$(7) \quad a_{it} = \theta S_{it} + \sum_k \zeta_k S_{-ikt}.$$

With (6),  $w_{ij} = \zeta_k d_{ij}^k$  whereas with (7),  $w_{ij} = \sum_k \zeta_k d_{ij}^k$ . More generally one can incorporate flexibility into the weighting matrix and estimate the associated distance function.

#### IV. MEASURING SPILLOVERS

In this section, we define our measures of spillovers in technology, product market, and geographic space. Several alternative distance or proximity measures have been used in the literature, and we build on and refine some of these.

IV(i). *Technology Spillovers*

We, like most researchers who examine technological proximity, use a measure that is due to Jaffe (1986).<sup>9</sup> Suppose that there are  $L$  technology classes,  $\ell = 1, \dots, L$ . Let  $F_{i\ell}^T$  be the fraction of firm  $i$ 's R&D activities (e.g., expenditures or patents) that are in class  $\ell$ . Firm  $i$ 's technology locational distribution is then  $F_i^T = (F_{i1}^T, F_{i2}^T, \dots, F_{iL}^T)$ . The measure of technology match between firms  $i$  and  $j$  is the uncentered correlation coefficient

$$(8) \quad w_{ij}^T = \frac{\sum_{\ell=1}^L F_{i\ell}^T F_{j\ell}^T}{\sqrt{[\sum_{\ell=1}^L (F_{i\ell}^T)^2][\sum_{\ell=1}^L (F_{j\ell}^T)^2]}}.$$

Notice that with (8) all classes are treated symmetrically — no class is ‘closer’ to any other. Furthermore, spillovers occur within but not across classes, or at least all ‘other’ technology markets are the same. In other words, there are no  $F_{i\ell}^T F_{jk}^T$  terms with  $\ell \neq k$ .<sup>10</sup>

Our technological spillover pool, SpillTech, is defined as

$$(9) \quad \text{SpillTech}_{it} = \sum_{j \neq i} w_{ij}^T S_{jt}.$$

IV(ii). *Horizontal Spillovers*

Most researchers use a discrete measure, same or different Standard Industrial Classification (SIC), that distinguishes between intra and interindustry spillovers (e.g. Levin and Reiss, 1988; Bernstein, 1988; Bernstein and Nadiri, 1989; Ornaghi, 2006). We use a continuous and more disaggregate measure that is employed by Bloom et al. (2013). In particular, suppose that there are  $M$  product markets,  $m = 1, \dots, M$ , where markets are proxied by SICs, and let  $F^P$  be defined similarly to  $F^T$ , i.e.,  $F_{im}^P$  is the fraction of  $i$ 's output that belongs to product market  $m$ . Then  $F_i^P$  is firm  $i$ 's product-market locational distribution, and  $w_{ij}^P$  is the uncentered correlation coefficient between the output distributions  $F_i^P$  and  $F_j^P$ . Notice that here, as with the technology measure, product markets are treated symmetrically — no market is ‘closer’ to any other — and spillovers occur within but not across markets.

Our horizontal spillover pool, SpillSIC is defined as

$$(10) \quad \text{SpillSIC}_{it} = \sum_{j \neq i} w_{ij}^P S_{jt}.$$

<sup>9</sup>The Jaffe measure suffers from limitations that are discussed by Jaffe as well as Cincera (2005).

<sup>10</sup>Bloom et al. (2013) consider extending this to measures that take into account the idea that some industries (and technological classes) may be closer to each other by using the Mahalanobis distance metric instead of the Jaffe style measure. They found similar qualitative results from the alternative measures.



IV(iii). *Geographic Spillovers*

Previous studies have tended to focus on the location of firms' headquarters as the basis for measuring geographic spillovers. Several measures have been used: i) a 0/1 variable that distinguishes if the firms' headquarters are located in a different or the same region (Adams and Jaffe, 1996; Eaton and Kortum, 1996; Orlando, 2004) and ii) a declining function of the Euclidean distance between the headquarters or the capital cities of the countries where the headquarters of firms  $i$  and  $j$  are located (Keller, 2002; Aldieri and Cincera, 2009).

It is not clear, however, that headquarters is the relevant locational variable. It is more likely that inventors communicate, and many inventors are employed in research labs. Moreover, unlike headquarters, a given firm might locate laboratories in many regions. Although the locations of headquarters and labs are correlated, this correlation is far from perfect, a fact that we document below.

In our empirical work we look at the locations of both headquarters and research labs. However, we do not have data on the geographic locations of labs. To circumvent this problem, we use patent data. Specifically, each patent gives the addresses of its inventors, and we construct geographic R&D locational distributions from those addresses.

We do this as follows. Suppose that there are  $K$  geographic regions,  $k = 1, \dots, K$ . Let  $F_{ik}^G$  be the fraction of firm  $i$ 's inventors that are located in region  $k$  and  $F_i^G$  be firm  $i$ 's geographic-region locational distribution. All of our geographic proximity weights are of the form

$$w_{ij}^G = \sum_k \sum_\ell f_{ijk\ell} C(d_{k\ell}), \quad i \neq j, \quad w_{ii}^G = 0, \quad C' \leq 0,$$

where  $d_{k\ell}$  is the Euclidean distance between regions  $k$  and  $\ell$  and  $f_{ijk\ell}$  is a function depending on the  $F_{ik}^G$ 's defined in equation (12) below. We assume that cross-regional spillovers are weakened, or at least not strengthened by geographic distance. Similar to our other two spillover variables, our geographic spillover pool, *SpillGeog*, is defined as a weighted sum of R&D stocks,

$$(11) \quad \text{SpillGeog}_{it} = \sum_{j \neq i} w_{ij}^G S_{jt}.$$

There are a number of attributes of our geographic measure that should be noted. First, unlike our product market and technology class measures, with our geographic measure spillovers can occur across regions but cross-regional spillovers are weighted by a declining function of geographic distance between regions. In addition, our measure specializes to headquarters, which is equivalent to considering firms that have only one lab or have all of their labs in the same region.

We distinguish between measures of *match* ( $f$ ) and measures of *Euclidean distance* ( $C$ ).

*Our measure of match.* We use the following function of the locational distributions,

$$(12) \quad f_{ijk\ell} = \frac{F_{ik}^G F_{j\ell}^G}{\sqrt{[\sum_{m=1}^K (F_{im}^G)^2][\sum_{m=1}^K (F_{jm}^G)^2]}}.$$

Note that, unlike the [Jaffe \(1986\)](#) measure, (12) allows for cross-regional spillovers ( $k \neq \ell$ ). We experimented with alternatives but ultimately settled on this one as it is a multidimensional relative of the uncentered correlation coefficient.<sup>11</sup>

*Measures of Euclidean distance.* We use two alternative functions of Euclidean distance. Specifically, we define the following functions  $C(\cdot)$  of the Euclidean distance,  $d_{k\ell}$ , between regions  $k$  and  $\ell$ ,

$$(13) \quad C_{\text{corr}}(d_{k\ell}) = \begin{cases} 1, & k = \ell, \\ 0, & k \neq \ell, \end{cases}$$

and

$$(14) \quad C_{\text{exp}}(d_{k\ell}) = \exp(-\alpha d_{k\ell}).$$

With the first function, the distance between regions does not matter. Indeed, all ‘other’ regions are the same regardless of distance. We use the subscript *corr* to indicate that this is the distance measure that is implicit when the uncentered correlation coefficient is calculated (i.e., our measures of technology and product-market spillovers). With the second measure, the effect of distance decays gradually, and the parameter  $\alpha$ , which determines the rate of decay, can be estimated or set exogenously. We use the subscript *exp* to denote exponential decay. The *exp* measure collapses into the one used by [Keller \(2002\)](#) when  $f(\cdot) \equiv 1$  and each firm has labs in only one region. Finally, when  $f(\cdot) \equiv 1$  and  $C = C_{\text{corr}}$  one obtains the popular o/1 measure. Our measures of geographic proximity therefore include those used by previous researchers as special cases.

To illustrate the construction of our geographic spillover measure, consider the example with three firms and four regions illustrated in Figure 1. Here,  $F_{23}^G = 1$  because all of firm 2’s inventors are located in region 3. Note that  $\sqrt{\sum_{m=1}^K (F_{im}^G)^2}$  (which is in the denominator of the right hand side of (12)) equals 0.6, 1, and 0.6 for  $i = 1, 2, 3$ , respectively. Thus,

$$f_{1324} = \frac{0.4 \times 0.4}{0.6 \times 0.6} = \frac{4}{9},$$

since 40% of firm 1’s inventors are in region 2 and 40% of firm 3’s inventors are in region 4. With  $C_{\text{corr}}$ ,  $f_{1324}$  is multiplied by zero because regions 1 and 4 are not the same, but with  $C_{\text{exp}}$ ,  $f_{1324}$  would receive a (presumably small) positive weight. To complete the example, note that using

<sup>11</sup>See [Lychagin et al. \(2010\)](#) for experiments with different specifications of match.

$C_{\text{corr}},$

$$\begin{cases} w_{12}^G = \frac{0.4 \times 0 + 0.4 \times 0 + 0.2 \times 1 + 0.0 \times 0.0}{0.6 \times 1} = \frac{1}{3}, \\ w_{13}^G = \frac{0.4 \times 0.2 + 0.4 \times 0.4 + 0.2 \times 0.0 + 0.0 \times 0.4}{0.6 \times 0.6} = \frac{2}{3}, \end{cases}$$

such that  $\text{SpillGeog}_{1t} = S_{2t}/3 + 2S_{3t}/3$ .<sup>12</sup>

## V. DATA

### V(i). Sources

We use two firm level data sources. Accounting data (sales, employment, capital, etc.) come from U.S. Compustat 1980–2000. Since we focus on manufacturing, we removed all firms whose primary industries are outside of the manufacturing sector (SIC codes 2000–3990). The data items available in Compustat, which are those reported in standard corporate income statements, are used to construct our measures of output and conventional inputs,  $Q$ ,  $K$ ,  $L$ , and  $M$ . We construct firm-specific price indices for sales, capital and materials using the NBER–CES Manufacturing Industry Database and information on firm activity by industry from the Compustat Segments data. We then use these price indices to deflate sales, book values of capital and material expenses to obtain  $Q$ ,  $K$ , and  $M$ .

The Compustat Segment data allow us to create breakdowns of each firm’s activities across product markets, which pinpoints the firm’s location in the product space. Specifically, firms’ sales are allocated to 308 3-digit industry codes.<sup>13</sup> On average each firm reports sales in approximately 5 different industry codes.

The Compustat firms were matched to the U.S. Patent and Trademark Office (USPTO) data from the NBER data archive. This archive contains detailed information on almost three million U.S. patents granted between January 1963 and December 1999 and all citations made to those patents between 1975 and 1999 (Hall et al., 2001; Jaffe et al., 1993). We kept only those firms that were assigned patents in the period of 1970–2000 leaving an unbalanced panel of 1383 firms. The USPTO allocates patents into 410 3-digit technology classes, such as class 042 (“Firearms”) and class 257 (“Solid state devices”), which is the breakdown that we use to construct technological proximity. On average each patenting firm owns 323 patents in 28 classes.

Inventor locations are taken from the addresses of the inventors of the patent, which are recorded at the city level. This is the standard measure of inventor location used inter alia by

<sup>12</sup>Please note that the fact that the weights add up to one is a coincidence here.

<sup>13</sup>The original Compustat Segments data often contain 4-digit product codes. However, in many cases 4-digit codes fail to represent a segment’s product as many segments are categorized under “catch-all” codes. Instead of relying on imprecise 4-digit codes we chose to aggregate segment data to the 3-digit SIC level. For instance, most segments producing industrial organic chemicals (3-digit SIC code 2860) are categorized under SIC code 2869 (“Industrial Organic Chemicals, Not Elsewhere Classified”). Other 4-digit codes from this group are 2861 (“Gum and Wood Chemicals”) and 2865 (“Cyclic Organic Crudes and Intermediates, and Organic Dyes and Pigments”). By using 3-digit SIC codes to pinpoint firm location in product space we pool these three types of industrial chemicals into one product category.

Jaffe and Trajtenberg (2005); Griffith et al. (2006), among many others. We feel that it is a more appropriate proxy for the location of the firm's R&D, and thus the potential for spillovers, than the headquarters of the firm, as it is a better indicator of where the key research was conducted. We allocate R&D activity into 2,070 geographic units, where a unit is a county. Because there are multiple patents, we are able to build up a picture of the location of the firm's R&D activity spatially. We do not use inventor information outside the United States. Since these are U.S. firms, most of their inventors are located inside the U.S. so we focus on within-U.S. interactions.

Each firm's own stock of knowledge,  $S$ , is constructed from R&D expenditure data as in equation (3) with the depreciation rate  $\delta$  set to 0.15 following Hall et al. (2005). We denote the horizontal product-market spillover measure defined by equation (10) SpillSIC and the technological spillover measure defined by equation (9) SpillTech. Finally, the measures of geographic proximity defined by equation (11) are denoted SpillGeog. Note, however, that SpillGeog varies according to the choice of function of geographic distance ( $C$ ) and whether we use inventor or headquarters locations.

#### V(ii). *Descriptive Statistics*

A more detailed description of the data is found in Appendix A, but we sketch some details in this section. Table I provides some basic descriptive statistics for the accounting and patenting data and the technology, product market, and geographic distance measures, SpillTech, SpillSIC, and SpillGeog. The sample firms are large (mean employment is over 11,000), but with much heterogeneity in size (measured by output, employment, or physical capital), R&D intensity, patenting activity, and location. The three spillover measures also differ widely across firms.

In order to distinguish between the effects of own R&D and technology, product market, and geographic spillovers econometrically we need independent variation in all of these variables. To gauge this we do two things.

First, we calculate and report in Table II raw correlation coefficients between pairs of measures of R&D stocks. To interpret Table II, consider the first column. The first row in that column contains the correlation coefficient between product and technology market spillover pools,  $\ln\text{SpillSIC}$  and  $\ln\text{SpillTech}$ , and the second between detrended and demeaned versions of those variables,  $\Delta\ln\text{SpillSIC}$  and  $\Delta\ln\text{SpillTech}$ .<sup>14</sup> Column (4) has the correlation between the firm's own R&D stock ( $\ln\text{OwnS}$ ) and its product market spillover ( $\ln\text{SpillSIC}$ ). In performing these calculations, we used the exponential measure of distance and inventor locations to construct SpillGeog. The correlations between the levels and the detrended and demeaned R&D stock variables are positive and significant at the 1% level. However, they are well below unity, implying substantial independent variation in each of the three measures. Finally, the third row contains raw covariances between spillover measures.

<sup>14</sup> $\Delta X$  is defined as the residual from a regression of  $X$  on firm fixed effects and time dummies, where  $X = \ln\text{OwnS}$ ,  $\ln\text{SpillSIC}$ ,  $\ln\text{SpillTech}$ , and  $\ln\text{SpillGeog}$ .

Second, we plot each pair of weights. Figures 2, 3, and 4, which show  $w^T$ ,  $w^P$ , and  $w^G$  plotted against each other, these figures reveal that the relationships between proximity measures are far from being perfectly correlated. For example, there are numerous pairs of firms that use a similar technology but are far apart in the product space. Furthermore, consistent with Table II, these figures show that the geographic spillover measures are less highly correlated with the other two proximity measures than those two are with each other.

We also investigate the relationship between the locations of inventors and headquarters in two ways. First, we calculate the aggregate distribution of inventor locations, which shows the aggregate fraction of inventors that are located at various distances from their headquarters. The quintiles of this distribution are reported in Table III, which shows that although many firms have all of their research activities located in close proximity to their headquarters, the activities of others are much more dispersed.

Second, we picked four firms that exhibit very different patterns. The information for those firms is summarized in Figure 5. The triangles in the graphs locate each firm's headquarters, whereas the circles show the distribution of its inventors. The first firm, Polaroid, has almost all of its research activities located in close proximity to its headquarters; the second, Motorola, has four concentrations of inventors, one of which is close to its headquarters; the third, Eaton (a power management company), has many concentrations of research activities. This dispersion is due to the fact that Eaton has expanded mainly through acquisitions; finally, the fourth, Union Carbide, has most of its R&D activities located in the Northeast where its headquarters used to be. However, when it relocated its headquarters to Texas it largely left its labs in place.

The geographic distribution of Motorola inventors demonstrates that we succeed in capturing the locations of major R&D labs. According to the company's website, Motorola had four U.S.-based development centers in 2009. Those centers are located in Schaumburg, IL, Phoenix, AZ, Plantation, FL, and Fort Worth, TX. This matches perfectly the four large clusters of Motorola inventors in Figure 5.

## VI. ECONOMETRIC METHODOLOGY

### VI(i). *The Model*

There are at least two approaches that can be used to assess productivity in our production-function framework. The first is to simply estimate equation (2) using firm and time period fixed effects to capture firm heterogeneity and aggregate shocks, respectively. To implement this approach one must specify a functional form for  $h$ , the function of conventional inputs, which for example could be Cobb–Douglas or translog. With this specification,  $a_{it}$  is a function of the firm's own knowledge stock and the spillover variables, as discussed earlier. Unfortunately, this approach suffers from an input endogeneity problem that we discuss below. However, there are methods of overcoming this problem with standard panel data econometric techniques. We experimented

with these formulations, and we discuss the results in Section VII(v). We only note here that the conclusions that can be drawn from the production–function and index–number estimations are essentially the same.

The second approach, which is the focus of our empirical work, involves constructing a productivity index, an index of outputs divided by an index of inputs, and regressing this index on own knowledge stocks and the spillover variables. There are many possible functional forms for productivity indices and some of them are exact for a flexible technology (e.g., the Tornqvist and the Fisher Ideal); in other words they provide a second–order approximation to an arbitrary technology. We use the former, which is exact for the translog.

Consider firm  $i$  and some comparison for  $i$ . The most standard comparisons are (i) to compare firm  $i$  in period  $t$  to firm  $i$  in period  $t - 1$ , (ii) to compare firm  $i$  in period  $t$  to firm  $k$  in period  $t$  (a unilateral comparison), or (iii) to compare firm  $i$  in period  $t$  to a hypothetical ‘average’ firm in period  $t$  (a multilateral comparison). We use a fourth and compare firm  $i$  in period  $t$  to a hypothetical ‘average’ firm  $i$ . This is very similar to the multilateral approach developed in Caves et al. (1982). However, whereas the averages there are taken over firms in period  $t$ , here they are taken over time for firm  $i$ . The reason is as follows. The Caves et al. method is usually used for country comparisons, where the production function is a country aggregate. Since macro production functions are arguably similar across countries, averaging across firms is not unreasonable in their context. In our case, however, we have firm data and firm technologies can be vastly different. The aggregation problem is therefore especially problematic. A possible problem with averaging over time is that there can be a trend in aggregate productivity.<sup>15</sup> The time–period fixed effects, however, solve this problem.

With our data we have only one output per firm,  $q_i$ , and therefore do not need to construct an output index. If each firm uses  $J$  inputs indexed by  $j$ , our Tornqvist productivity index is<sup>16</sup>

$$(15) \quad \text{TFP}_{it} = (q_{it} - \bar{q}_i) - \sum_j \frac{1}{2} (s_{ijt} + \bar{s}_{ij}) (x_{ijt} - \bar{x}_{ij}),$$

where  $s_j$  is factor  $j$ ’s share of costs and bars denote firm–specific time averages.<sup>17</sup> This index is regressed on our spillover measures, own R&D, and firm and time–period fixed effects.

There are several properties of our index number approach that are worth noting. First, the production function parameters need not be estimated. Second, the input endogeneity problem is solved by moving the inputs to the left hand side of the equation. Finally, the translog parameters can differ by firm up to the second order. In particular, this means that if the technology were Cobb–Douglas, the functional form that is used in most of the R&D spillover literature, all parameters

<sup>15</sup>Note that, since we estimate rates of change, there is no obvious trend.

<sup>16</sup>Recall that lower case variables are in logarithms.

<sup>17</sup>We assume constant returns to scale and competitive markups, which imply that it doesn’t matter if we use cost or revenue shares. The assumption of constant returns is consistent with the production function estimates that are reported in Section VII(v). The zero markup assumption is more problematic since the two assumptions combined imply that firm sizes are indeterminate.

would be firm specific.

We include a firm's own stock of knowledge  $\ln\text{OwnS}$ , the horizontal product-market spillover measure  $\ln\text{SpillSIC}$ , and the technological spillover measure  $\ln\text{SpillTech}$  as explanatory variables, where variables that are preceded by  $\ln$  are in natural logarithms. The presence of own R&D on the right hand side may appear puzzling because in the firm accounts R&D will be captured by labor, capital or materials. Thus, it will already be captured in the TFP measure insofar as it is a conventional factor of production. We include it in our estimating equation first because this makes our estimates more comparable with the existing literature but second because R&D is related to the quality of standard inputs like labor and capital.<sup>18,19</sup> Most equations that are reported below contain all of these variables. In addition, they contain one or more measures of geographic proximity,  $\ln\text{SpillGeog}$ , that differ according to the choice of geographic distance function (C) and whether an inventor or headquarters measure is used.

Our index-number estimating equation is thus<sup>20</sup>

$$(16) \quad y_{it} = \beta_0 + (z_{it} - \bar{z}_i)\beta + \omega_t + u_{it}, \quad i = 1, \dots, n_t, \quad t = t_{i0}, \dots, T_i,$$

with  $y_{it} = \text{TFP}_{it}$  and

$$(17) \quad z_{it} = [\ln\text{OwnS}_{it}, \ln\text{SpillSIC}_{it}, \ln\text{SpillTech}_{it}, \ln\text{SpillGeog}_{it}],$$

where  $\bar{z}_i$  denotes a within firm average. The number of firms differs by period and the years in which a firm is observed differ by firm because the panel is unbalanced.

## VI(ii). *Econometric Issues*

*Identification.* The potential bias in OLS estimates of production functions has long been recognized (see e.g., [Marschak and Andrews Jr, 1944](#)). This bias results from the possible correlation between input levels and firm level productivity shocks. Specifically, when firms experience a large productivity shock, they might respond by using more inputs. Applied economists have devised alternatives to OLS that attempt to circumvent this problem. Most use either a variant of the method developed by [Olley and Pakes \(1996\)](#) and extended by [Levinsohn and Petrin \(2003\)](#); [Gandhi et al. \(2013\)](#); [De Loecker \(2011\)](#) or the GMM methods proposed by [Arellano and Bond \(1991\)](#); [Arellano and Bover \(1995\)](#); [Blundell and Bond \(2000\)](#). We have chosen to focus on GMM approaches because it is not straightforward to introduce endogenous R&D into Olley–Pakes

<sup>18</sup>In addition, the econometric production function specifications include  $\ln K$ ,  $\ln L$ , and  $\ln M$ .

<sup>19</sup>Some of our firms do not perform R&D and thus have zero stocks; observations with zero R&D stocks account for 8% of the sample. To overcome this problem, we follow [Klette \(1996\)](#) and set  $\ln\text{OwnS} = 0$  when  $S = 0$  and include a dummy variable that equals one when this occurs. We experimented with several other specifications for the  $\text{OwnS}$  variable and found that our basic conclusions were robust to these changes.

<sup>20</sup>Our demeaned specification produces estimates that are almost identical to one that includes firm fixed effects and levels of the dependent and explanatory variables (i.e., without demeaning).



approaches. (see, e.g., [Akerberg et al., 2007](#), as well as [Buettner, 2004](#)).<sup>21</sup>

With the index–number specifications, the conventional inputs do not pose an endogeneity problem because they appear on the left hand side of the estimating equation.<sup>22</sup> Nevertheless, the own stocks of knowledge are potentially endogenous.<sup>23</sup> We therefore also consider the [Arellano and Bond \(1991\)](#) correction for endogeneity to  $\ln\text{OwnS}$ .

Identification of (16) relies on within–firm variation of the spillover measures. We have shown that, in the data, there is substantial independent variation in the measures. However, much of that variation is cross sectional. We therefore assess time–series variation within a firm, first with an example and second with a figure.

Suppose that there are three firms, the firm of interest  $i$ , a second firm  $g$  that is in close geographic (but not technological) proximity to  $i$ , and a third  $t$  that is in close technological (but not geographic) proximity. When firm  $g$  varies its R&D expenditures,  $\text{SpillGeog}_i$  will vary while  $\text{SpillTech}_i$  remains constant. The opposite will be true when firm  $t$  changes its R&D efforts. Differences in the weights therefore lead to independent variation in  $i$ 's spillovers from the three sources. The firm–specific spillover weights are (in our baseline estimates) held fixed over time, but they are specific to every dyad of firms. Firms are subject to differential shocks in R&D over time, and the effect of such a change in R&D will be heterogeneous across firms depending on the distance metric (e.g. geographical vs. patent class). So identification comes from the interaction between a time invariant characteristic and a time-varying shock.

While the example shows that, in theory, independent variation exists, we must also show that it occurs in practice. Figure 6 assesses this issue. The figure shows the distribution across firms of within–firm correlation between detrended spillover pairs. The first part of the figure, which assesses within firm correlation between  $\ln\text{SpillGeog}$  and  $\ln\text{SpillTech}$ , shows that, although correlation is high for some firms, there is a substantial number of firms for which correlation is low or negative. The same is true for correlation between geographic and product market spillovers. In contrast, the correlation between technological and product market spillovers is higher, confirming our earlier findings.

<sup>21</sup> Essentially, the [Arellano and Bond](#) approach assumes that serial correlation of the error term (after controlling for fixed effects) is of finite order. In the simplest case of no serial correlation this implies that levels dated  $t - 2$  and earlier are valid instruments. If there is some first order correlation (e.g. MA(1)) we can use  $t - 3$  dated instruments. One can also use the additional moments for the levels equations suggested by [Arellano and Bover \(1995\)](#) and first used in a production function context by [Blundell and Bond \(2000\)](#). The GMM approach does have issues, however. First the instruments may have weak power in the first stage. Second, the construction of our TFP index uses past and future realizations of shocks to inputs and outputs, so will likely violate the assumption of no serial correlation of  $u_{it}$  in (16). There is no one perfect way to estimate the TFP equations, so we consider it useful to contrast OLS and GMM to see whether our conclusions are substantially robust.

<sup>22</sup> The index–number and production–function approaches are based on different assumptions about the way in which conventional inputs adjust. In particular, the first is based on the assumption that inputs adjust instantly whereas the second requires that all inputs adjust with a lag.

<sup>23</sup> It is more plausible to take rival knowledge stocks as exogenous for three reasons. First, in a competitive model, R&D levels are not chosen jointly. Second, we follow the spillover literature, as well as the differentiated–products literature, in assuming that a firm's position in space, whether it be geographical, technological, or product–characteristic space, is exogenous. See, e.g., [Keller \(2002\)](#); [Berry \(1994\)](#) for the two literatures. Third, [Bloom et al. \(2013\)](#) consider instrumenting spillovers with R&D tax credits and found that their OLS and IV results were similar.



*Estimation.* We use two different methods to estimate the unknown coefficients in (16): OLS and GMM. The OLS estimator here is essentially equivalent to the standard fixed effects estimator, i.e., the OLS estimator after subtracting out the means (over time) for each firm. For GMM we consider both a static and a dynamic version.

For the static versions of GMM we use the methodology of [Arellano and Bond \(1991\)](#) with sufficiently lagged  $x_{it}$ 's as instruments. This presumes that such instruments are orthogonal to current and (one-period) lagged errors. This assumption is substantially weaker than the assumption of strict exogeneity of the instruments even if the errors are serially dependent.

The dynamic version of the model is similar to that of [Blundell and Bond \(2000\)](#). Define TFP in levels,<sup>24</sup>

$$(18) \quad \text{TFPL}_{it} = q_{it} - \sum_j \frac{(s_{ijt} + \bar{s}_{ij})x_{it}}{2},$$

and consider an equation in levels with firm fixed effects,

$$(19) \quad y_{it}^\ell = z_{it}\beta + v_i + \omega_t + u_{it},$$

where  $y^\ell = \text{TFPL}$ . If the error term follows the AR(1) process

$$(20) \quad u_{it} = \rho u_{it-1} + \varepsilon_{it},$$

where  $\varepsilon_{it}$  is serially independent, one can substitute (19) into equation (20) to obtain

$$(21) \quad y_{it}^\ell = \pi_1 y_{it-1}^\ell + z_{it}\pi_2 + z_{it-1}\pi_3 + v_i^* + \omega_t^* + \varepsilon_{it},$$

where  $v_i^* = (1 - \rho)v_i$ ,  $\omega_t^* = \omega_t - \rho\omega_{t-1}$ ,  $\pi_1 = \rho$ , and  $\pi_2, \pi_3$  satisfy the common factor restriction

$$(22) \quad \beta = \pi_2 = -\frac{\pi_3}{\pi_1}.$$

We estimate the unrestricted model of equation (21) and then impose (22) by minimum distance methods.

Finally, the standard errors that are shown in all tables are clustered at the firm level.

## VII. TFP INDEX NUMBER ESTIMATES

The Tables IV–X contain a number of specifications of our index number equation with knowledge spillovers, all of which include year fixed effects. We begin with the OLS estimates.

---

<sup>24</sup>We formulate the GMM equation in levels with firm fixed effects because the [Arellano and Bond](#) estimation differences the data to remove those effects, which is the same as demeaning.

## VII(i). OLS

*Single spillover measures.* Table IV, which assesses each spillover measure in isolation as well as all three jointly, shows that when only one measure is considered, the coefficient of that measure is positive and significant at conventional levels. However, the coefficients of  $\ln\text{SpillTech}$  and  $\ln\text{SpillGeog}$ <sup>25</sup> are much larger than the coefficient of  $\ln\text{SpillSIC}$ , even though their standard deviations are of similar size. This remains true when all three measures are included in the same equation. Moreover, all spillover coefficients are smaller in the fourth specification, which is due to the fact that the spillover variables are positively correlated and thus the single-measure coefficients are biased. This clearly implies the need to consider a multi-space model as we do.

With some specifications, the coefficient of  $\ln\text{OwnS}$ , although positive, is not significant, a regularity that persists in most of the later tables. When we drop the observations with  $\text{OwnS} = 0$ , or when we use other measures of the R&D stock, the coefficient of that variable becomes positive and significant. For example, when  $\sqrt{\text{OwnS}}$  is substituted for  $\ln\text{OwnS}$ ,<sup>26</sup> its coefficient is significant. However, in the interest of full disclosure we decided to report specifications with  $\ln\text{OwnS}$  because its construction is consistent with previous literature. In addition, our research does not focus on the magnitude or significance of the coefficient of that variable.

*Multiple spillover measures.* Table V contains further estimates of the index number equation. As indicated in the table, the first three specifications (columns 1–3) are based on the locations of inventors using our new geographic spillover measure. The next two (columns 4 and 5) are based on the locations of headquarters as has been customary in previous work, and the last two columns (6 and 7) contain both inventor and headquarters measures. Furthermore, specifications 1 and 2 for inventors (4 and 5 for headquarters and 6 and 7 for both inventors and headquarters) are distinguished by the specification of the distance function  $C$ . The first of each pair is based on the exponentially decaying distance function (Exp) and the second on the  $o/1$  distance function (Corr). Finally, column (3) shows a specification that includes both the exponential and the  $o/1$  distance function for inventors.

First consider the top half of the table, which contains the variables that are common across equations. It shows that the signs, magnitudes, and significance of the coefficients of those variables are stable across specifications of the geographic distance function. Furthermore, all coefficients are similar to those found in column (5) of Table IV, which is column (1) in Table V.

To assess geographic spillovers in more detail, one must examine the bottom half of Table V. First, compare inventor-based measures (denoted INV) to headquarter-based measures (denoted Head). These specifications are found in columns (1) and (4) and (2) and (5). With both comparisons, the coefficients of the inventor measures are much larger than those of headquarters measures.

<sup>25</sup>In this table, the  $\text{SpillGeog}$  variable is constructed using the inventor measure of match and the exponential distance function.

<sup>26</sup>Taking the square root of a persistent variable to overcome the zero-value problem and to capture diminishing returns is common in the advertising literature. See, e.g., Gasmi et al. (2005), Slade (1995); Kadiyali (1996).

Finally, when both inventor and headquarters measures are included (columns 6 and 7) the coefficients of the inventor variable are larger and with the second comparison, the headquarters coefficient is insignificant. These findings imply that the elasticity of TFP with respect to our new inventor-based measure is at least twice as large as,<sup>27</sup> and generally more significant than, that for the headquarters measure.

Second, compare the exponentially decaying geographic distance function (denoted Exp) to the zero/one measure (denoted Corr). With the Corr measure, the region in which a firm performs R&D matters. Indeed, a firm that has a lab (headquarters) in a particular location benefits from the R&D activities of rival firms in that location. In contrast, with the Exp measure, not only does the region in which one locates matter but distance between regions matters as well. In particular, firms benefit from R&D performed in more distant locations but at a rate that decays with geographic distance. The comparisons between Exp and Corr, which can be found in columns (1) and (2), (4) and (5), and (6) and (7), show that the coefficients of the exponential measures are always much larger and more significant than those of the 0/1 measure. Finally, when measures based on both distance functions are included in the same equation (16), the coefficient of the Corr measure is insignificant. One can therefore conclude that ignoring inter regional spillovers and the distance between regions results in an underestimate of geographic spillovers.

Table V thus reveals two regularities that persist in alternative specifications. First, although match — having local R&D distributions that are similar — by itself is important, it is clear that considering both match and Euclidean distance is associated with estimates of spillover effects that are both larger and more significant than when distance is ignored. Second, geographic distance effects are larger when inventor rather than headquarter locations are chosen. Our use of the patent data therefore appears to result in an improved geographic measure.

To demonstrate the magnitude of geographic spillovers we perform the following thought experiment.<sup>28</sup> We track the total factor productivity (TFP) of a hypothetical firm, whose inventors are located in Los Angeles County (“incumbent LA firm”). We assume that another hypothetical firm has a lab in the same county with an R&D stock of \$500m (1987 prices). We then relocate this lab farther and farther away from LA. As the lab moves to an infinite distance, its contribution to the LA incumbent’s geographic spillover variable declines to zero, while the product market and the technology spillovers are left unaffected. Having such a lab in LA raises the incumbent LA firm’s TFP by 0.84%, whereas having the mobile R&D lab in San Diego would raise TFP by only 0.43%. If the lab were moved to San Francisco, its contribution would decline further to 0.13%. Having the lab as far as Seattle would only raise the LA-based firm’s TFP by 0.005% (essentially zero). This is a substantial effect and could rationalize why cities and states compete so eagerly for the location of R&D activities (Wilson, 2009).

Finally, we assess the relative magnitudes of spillover effects across firms by looking at the

<sup>27</sup>This is true because, in our setting, the ratio of the coefficients of two spillover measures equals the ratio of the comparable elasticities.

<sup>28</sup>This thought experiment is based on the estimates from the last column of Table IV.

implied impact on TFP of a one standard deviation change in each measure.<sup>29</sup> All comparisons are relative to *lnSpillGeog* calculated using exponential decay with distance and inventor locations. According to this calculation, the effect of geographic spillovers is 8% higher than that of technology spillovers and 367% higher than that of product market spillovers. These comparisons should be interpreted cautiously, however, since our measure of geographic spillovers is more sophisticated than the other two. Comparing different geographic measures, we find that the inventor measure based on exponential decay is 32% higher than the headquarters measure that is based on exponential decay and is 235% higher than the inventor measure that uses a  $0/1$  distance function.

#### VII(ii). GMM

An obvious concern with the estimation strategy is that, even after controlling for heterogeneity at the firm and sectoral level, the own knowledge stocks are endogenous. Column (1) in Table VI presents GMM estimates of the static index number formulation (an IV estimate of column (1) in Table V) using Arellano and Bond instruments. The results are consistent with our previous estimates: geographically and technologically-based knowledge spillovers are positive and significant, albeit with coefficients that are smaller than the OLS estimates.

To evaluate instrument strength we conducted both standard first stage F tests and the minimum eigenvalue test proposed in Cragg and Donald (1993), the results of which were statistically acceptable.<sup>30</sup> Further, the serial correlation tests at the base of column (1) suggest no problems (we expect first order serial correlation due to the MA(1) nature of the first differenced errors). However, the Hansen test rejects the null of instrument validity which could indicate misspecification. If we assume that there is an AR(1) error term in the TFP equation as in (19) and (20), this generates a model that includes a lagged dependent variable and lags of all explanatory variables in the first equation, which we show in column (2).

As outlined in Section VI(ii), the structure of the model implies common factor (COMFAC) restrictions on the coefficients in column (2), which we impose in column (3). The qualitative results from this COMFAC model are consistent with the findings in column (1) and elsewhere in the paper. In particular, there is a significant effect of the geographically-based R&D spillover term in addition to the standard technological spillover term. As in column (1), however, there remain diagnostic problems. The Hansen test is slightly better (but still rejects at the 10% level), there is

<sup>29</sup>In other words, we multiply the coefficient of a spillover variable times the standard deviation of that variable.

<sup>30</sup>The individual F statistic values were greater than 11 so they met the Staiger and Stock (1997) cutoff. However, nonidentification and weak identification can occur even if some of the coefficients in the first stage are nonzero. We therefore used the (F-statistic implementation of the) minimum eigenvalue test, which verifies that the rank condition required by IV-type methods is satisfied. The statistic value was 4.5 for 106 numerator degrees of freedom. This suggests that the null of nonidentification is soundly rejected (with a p-value equal to zero up to at least six decimal places). Stock and Yogo (2005) propose cutoff values that are much greater than the 5% critical values of the minimum eigenvalue test since it is possible to construct examples in which GMM type estimators perform poorly if the first stage F-statistics fall below the thresholds they compute. The cutoff values listed in their paper depend on the estimation method used, none of which corresponds exactly to ours. Our F statistic value exceeds the cutoff value listed in Stock and Yogo (2005) for two out of the three methods listed plus it should be noted that these cutoffs are based on the worst case, not the typical, scenario.

now evidence of second order serial correlation, and the COMFAC restrictions are rejected.<sup>31</sup>

In summary, the GMM results broadly support the earlier findings. In particular, the positive and significant effect of the spillover terms is invariant to many experiments with different specifications. Although no one specification is ideal, we feel that our key conclusions are robust.

### VII(iii). *Alternative TFP Index Number Specifications*

It seems clear that location and distance are important determinants of R&D spillovers. Several questions, however, remain. First we ask why distance matters and second we ask if, even though distance is not dead, it might be in the process of dying. Finally we revisit the endogeneity issue.

*Interactions between spillovers.* Geographic distance could be an important determinant of R&D spillovers for the reasons that we state in the introduction; that is, importance could be due to the formation of social networks, R&D agglomeration effects, and within-region labor mobility. However, there are alternative hypotheses. From the outset, the literature has had difficulty distinguishing between ‘pure’ distance effects and distance effects that reflect differences in local and distant technologies. We now try to dig deeper into the reasons why knowledge flows fade with distance. In particular, knowledge flows might diminish because face-to-face meetings, second and third-hand information flows, and R&D agglomeration effects might decline as distance increases. We call this the ‘declining-contact-with-distance’ hypothesis. Alternatively, knowledge flows might decline with distance because nearby knowledge is on average more similar and more relevant — through industry clustering — than distant knowledge. We call this the ‘decreasing-relevance-with-distance’ hypothesis.

Although we have conditioned our estimates of geographic spillovers on measures of technology and product market spillovers, we have not yet included interactions that would allow us to distinguish between the two hypotheses. Table VII contains specifications of our index-number TFP equation that include interactions of  $\ln\text{SpillGeog}$  with  $\ln\text{SpillSIC}$  ( $\text{SICGeog}$ ) and  $\ln\text{SpillTech}$  ( $\text{TechGeog}$ ). If the declining-contact-with-distance hypothesis dominates, the coefficients of the interaction variables should be insignificant and that of the geographic spillover variable should remain unchanged. If, in contrast, the decreasing-relevance-with-distance hypothesis dominates, the opposite should be true.

Table VII shows specifications with each interaction variable included by itself as well as one that contains both interactions. One can see that, with all three specifications, the coefficients of the interaction variables are insignificant. Furthermore, the coefficient of  $\ln\text{SpillGeog}$  is stable. Finally, when  $\text{SICGeog}$  is included, the coefficient of  $\ln\text{SpillSIC}$  loses its significance. Since the

<sup>31</sup>We also experimented with using further “levels” moments that follow if we make assumptions over the initial conditions (essentially mean stationarity) as suggested by [Arellano and Bover \(1995\)](#); [Blundell and Bond \(1998\)](#). Unfortunately these additional levels moments did not appear to be valid in our context. In particular, the coefficient on the lagged dependent variable in the COMFAC model was one (greater than one in some specifications) and its t statistic was 171 (several hundred in some specifications). Moreover, many of the other coefficients were negative.

magnitude of this coefficient is stable, the inclusion of SICGeog may simply introduce too much noise. Alternatively, insignificance could be taken as evidence that estimated product market spillovers are to some extent due to geographic clustering of firms that produce similar products (e.g., automakers in Detroit).<sup>32</sup> Regardless of how one interprets the coefficients of  $\ln\text{SpillSIC}$ , however, the results in Table VII support the declining–contact–with–distance hypothesis.

*Is distance dying?* Our data span the 1980 – 2000 period, a period when the use of computers, the internet, and wireless technology was growing at a fast pace. It is therefore possible that, even if distance is not dead, its importance is fading with time. We call this the ‘distance–is–dying’ hypothesis.

To determine if geographic spillover effects became less important in later years, or, in other words, if the areas over which ideas are easily communicated became larger, we broke the sample into two periods, the 80s and the 90s. We then recalculated the spillover weights,  $w_{ijt}^G$ , based on inventor locations in the two time periods and recalculated  $\text{SpillGeog}$  based on the period weights. We then estimated the half life of the geographic spillover effect — the distance at which one half of the spillover is dissipated — for the entire period as well as for each subperiod.<sup>33</sup>

Table VIII, which contains two specifications of the TFP equation with estimated half lives, one for the entire period and one in which the inventor weights and half lives differ by subperiod, shows that the estimated half lives are remarkably similar across periods. We therefore conclude that distance is not dying, or at least it was not showing signs of mortality at the end of the 20th century.

*A four–factor TFP index.* Although our dynamic GMM estimations are supportive of our main conclusions, diagnostic problems remain. In particular, the COMFAC restrictions are rejected. We therefore return to the endogeneity issue for our final index number specification. The potential endogeneity problem arises because the stock of own R&D,  $\ln\text{OwnS}$ , appears on the right–hand side of our estimating equation. To circumvent this problem we calculated a four factor TFP index that includes not only capital, labor, and intermediate inputs, but also the stock of own R&D.<sup>34</sup> Table IX, which is comparable to Table IV with share–weighted  $\ln\text{OwnS}$  subtracted from the left hand side, contains OLS estimates of the four–factor TFP equation. One can see that the coefficients and t statistics in the two tables are very similar. This evidence, combined with our earlier GMM estimations, leads us to conclude that endogeneity cannot overturn our main conclusions.

<sup>32</sup>The insignificance of  $\ln\text{SpillSIC}$ , when  $\text{SICGeog}$  is included is also consistent with the model in Bloom et al. (2013), where R&D by product market rivals affects market value negatively due to a business–stealing effect, but has no effect on total factor productivity, as it does not alter technological capabilities conditional on own R&D.

<sup>33</sup>Note that the half life is functionally related to  $\alpha$  in equation 14.

<sup>34</sup>The share of R&D is calculated as a ratio of R&D expenditures to sales. It is also assumed that 50% of R&D expenses account for labor, and another 50% account for intermediate materials. We adjust the respective expenditures to avoid double accounting.



VII(iv). *Alternative Explanations*

As we have demonstrated, lab and headquarters locations are correlated, if only imperfectly. It is also true that plant and headquarters locations are imperfectly correlated, which raises the possibility that the geographic spillover effects that we are finding are spurious and are in fact due to colocation of firms' labs and plants. In other words, it is possible that the TFP effects that we are attributing to R&D spillovers between labs are really agglomeration effects associated with proximity of production facilities. In this subsection, we explore this and other alternative explanations for our results.

Agglomeration spillovers could confound our findings through the following channel. Suppose that firm  $j$  performs R&D that raises its productivity (its output, holding inputs constant). An increase in  $j$ 's output would cause  $i$ 's output to rise through the agglomeration effect. If agglomeration effects are strong and colocation is common, the R&D spillovers that we have measured could be spurious.

More formally, in the presence of agglomeration effects, the production function takes the form

$$(23) \quad q_i = f(s_j, s_i, q_j, x_i),$$

where lower case letters denote logarithms,  $s_j$  is shorthand for the pool of geographic-distance weighted rival stocks, and  $x_i$  is a vector of conventional inputs. We include distance-weighted rival output,  $q_j$ , to capture the agglomeration effect.

Differentiating equation (23) holding inputs constant we obtain

$$(24) \quad \frac{dq_i}{ds_j} = \frac{\partial q_i}{\partial s_j} + \frac{\partial q_i}{\partial s_i} \frac{ds_i}{ds_j} + \frac{\partial q_i}{\partial q_j} \frac{dq_j}{ds_j}.$$

The first term in equation (24) is the spillover effect that we have measured. In the second term,  $ds_i/ds_j$  is the response of own R&D to rival R&D in the current period. Since the increments to the R&D stocks are chosen simultaneously, this term will be zero in either a competitive equilibrium or a Nash equilibrium of a static R&D game.<sup>35</sup> This leaves the third term, which is the potential agglomeration bias that we have assumed to be zero. We don't have data on the locations of firms' plants and thus can not calculate the agglomeration effect,  $\partial q_i/\partial q_j$ . However, we know that this derivative must be less than one. A value for the second part of the third term can be found in Table IV. In particular, the coefficient of  $\ln \text{OwnS}$  in the first equation in that table, which is 0.05, is our estimate of  $dq_j/ds_j$ .<sup>36</sup>

These calculations imply that, although an agglomeration bias might be present, it is small.

<sup>35</sup>We assume that R&D markets are competitive. With a static Nash equilibrium, firms would simultaneously choose inputs and R&D expenditures, and those choices, plus the spillovers and  $s_{it-1}$  and  $s_{jt-1}$ , would determine their outputs.

<sup>36</sup>Since we do not include the spillover variables in the first equation in Table IV, this coefficient captures the indirect effects of those variables. In other words, it is the total effect,  $dq_j/ds_j$ , not the partial effect,  $\partial q_j/\partial s_j$ .

Indeed, it is less than 0.05,<sup>37</sup> whereas the spillover effect that we measure is large, approximately 0.8. After subtracting the potential agglomeration bias from our estimate, we still obtain a substantial geographic spillover effect of 0.75.

It would nevertheless be interesting to test the agglomeration hypothesis more formally. This could be accomplished by constructing an output based measure — a weighted sum of other firms' outputs calculated using weights that are based on the geographic proximity of firms' plants, not labs — and including that measure in our regressions. However, we do not have data on the locations of firms' plants. We therefore construct an imperfect proxy based on the geographic R&D location weights,  $w_{ij}^G$ ,

$$(25) \quad \text{Spill}X_{it} = \sum_{j \neq i} w_{ij}^G X_{jt},$$

where  $X$  is some measure of a firm's production or size. This measure is exact if production and R&D locations are perfectly correlated. To the extent that the two differ, however, the usefulness of the proxy declines.

It is not clear how to choose the best measure of firm size. Indeed, the most obvious candidate, output or  $Q$ , is likely to be highly endogenous. We have therefore chosen two alternative size measures, a firm's physical capital  $K$ , which is least likely to be endogenous, and its employment  $L$ . Table X contains specifications that include these size measures. Column 1 is our baseline specification (column 1 in Table V) whereas the size based measures  $\ln\text{Spill}K$  and  $\ln\text{Spill}L$  have been added to columns 2 and 3, respectively. The table shows that, with either measure of size, the coefficient of the R&D spillover variable is significant at the 5% level but that of the size spillover is not.

Finally, it is possible that firms chose to locate their research labs in regions where other labs are present, in which case locations would be endogenous. However, we are interested in the short-run effects of R&D, conditional on the long-run choice of location.<sup>38</sup> Our findings show that, if firms chose to cluster their labs, those choices were good ones.

#### VII(v). *Production Function Estimates*

We emphasize our index number specifications because they are more flexible and because they do not suffer from the input-endogeneity problem. Nevertheless, it is natural to ask if estimates of R&D spillovers based on an econometric model of the production function are similar or if some of our conclusions are reversed when we adopt that approach. We therefore report some production function specifications. For our production function estimations, we use a Cobb–Douglas functional form, as is standard in this literature, and include both firm and year fixed

<sup>37</sup> The comparable value for the production function coefficient of  $\ln\text{Own}S$  is 0.035, which is even smaller.

<sup>38</sup> Our analysis is comparable to the numerous studies of short run output or price choices conditional on location in product market or product characteristic space (see, e.g., Berry et al., 1995).



effects.

Table XI is comparable to Table V. In contrast to that table, however, the dependent variable here is  $\ln Q$  and the conventional inputs,  $\ln K$ ,  $\ln L$ , and  $\ln M$ , are included on the right-hand side. One can see that the results are very similar to those reported earlier. In particular, all of our main conclusions continue to hold. Furthermore, the coefficients of the conventional inputs are stable across specifications. Finally, with these specifications, in contrast to the TFP versions, one can assess economies of scale. Table XI shows that the sum of the coefficients of the conventional inputs is very close to, but slightly less than, one, which indicates that constant returns to scale is not a bad approximation.

We have estimated the production function equivalents of the other index number tables and have found that our conclusions are robust. To save on space, however, we do not report these specifications.<sup>39</sup>

## VIII. CONCLUSIONS

A number of conclusions can be drawn from our study. First and foremost, geography matters. Indeed, intraregional spillovers are significant, sizable, and economically important, even after conditioning on technological and product-market spillovers. In other words, a lab's location is an important determinant of a firm's productivity. This finding is consistent with an emerging literature on plant, not lab, location, which suggests important local production spillovers (Greenstone et al., 2010). It is also consistent with the sizable literature on agglomeration economies, as summarized in Rosenthal and Strange (2003). We conclude that social learning and capitalization of complementarities among firms' research activities can be important factors for economic growth.

Our finding is also consistent with the idea that, when skilled workers change jobs, they are apt to move to a different firm in the same geographic region and they take their knowledge with them. For example, Fallick et al. (2006) and Freeman (2008) provide evidence that U.S. skilled employees move rapidly among firms in high tech clusters in the computer and software publishing industries. In addition, Almeida and Kogut (1999) investigate interfirm intraregional mobility of U.S. engineers and conclude that the flow of knowledge is embedded in regional labor networks. Job hopping within regions is also common in other countries. For example, Combes and Duranton (2006) find that, when French skilled workers change jobs, about 75% stay in the same region. Finally, even when an inventor leaves a region, due to the persistence of social networks, the region can still benefit from the inventor's ideas. Indeed, Agrawal et al. (2006) confirm this hypothesis and find that the effect is especially important across research fields.

Second, interregional spillovers matter. More precisely, a geographic distance function that allows spillovers to decay gradually as regions become farther apart outperforms a specification

---

<sup>39</sup>All production function tables are available from the authors upon request.

that constrains spillovers to occur only within regions (i.e., a zero/one distance function that indicates that two regions are the same or different).

Third, estimated geographic spillover effects are larger when the distribution of each firm's inventors is used as a measure of R&D location rather than the location of its headquarters. Our use of the patent data to create spatial distributions of the location of firms' research activities therefore appears to be a worthwhile extension.

Finally, not only is distance not dead, but it does not appear to be dying, as witnessed by the fact that, even though our data span a period of rapid growth of computerization of information and instant access to ideas, estimated geographic spillovers were just as local in the 1990s as they were in the 1980s.

Turning to technological spillovers, we have experimented less with different measures of this important variable. Nevertheless, as with geographic spillovers, we find that technological spillovers are significant, sizeable, and economically important. In fact, in most specifications, the magnitude of the coefficients of this measure are comparable to those of the geographic measures.<sup>40</sup>

The picture with respect to product-market R&D spillovers is somewhat different. Indeed, although the coefficient of this variable is positive in most specifications, it is always smaller and economically less important. This finding is consistent with the model outlined in [Bloom et al. \(2013\)](#), where R&D by product market rivals affects market value negatively due to a business-stealing effect, but should have no effect on total factor productivity, as it does not alter technological capabilities conditional on own R&D.

What have we learned from a policy point of view? We can conclude that since estimated spillover effects are large, there is a sizable public-good aspect to R&D activity. In the absence of public policy to rectify this externality, we might therefore expect to see underinvestment in R&D. Nevertheless, as there are also costs associated with intervention, more research would be required before one could gauge whether the current amounts of R&D subsidy are optimal or not. In addition, we find that geographic spillovers are moderately local, which is bad news for regional convergence. This could explain, however, why federal policymakers subsidize the R&D activities of domestic firms and why state officials invest substantial sums in tax incentives to attract R&D labs to their states ([Wilson, 2009](#)).

Although we have used specifications of proximity measures that allow for cross-regional spillovers only for geographic regions, the same could be done for product and technology markets. To illustrate, consider product markets. It is clear that the production of cars is much closer to that of trucks and buses than it is to that of breakfast cereals and ladies apparel. However, the zero/one distance function that is implicitly assumed in research to date does not allow for intermediate cases (i.e., two activities are in different but close product markets).<sup>41</sup> It might therefore be fruitful

<sup>40</sup>Note, however, that we have given the geographic measure the benefit of the doubt, since we pay more attention to its construction.

<sup>41</sup>This extension is similar to the methods developed in [Pinkse and Slade \(2004\)](#) to model differentiated products.

to modify the product and technology proximity measures to incorporate nontrivial distance functions. Moreover, the construction of SICs and technology classes, which involves classification into  $n$ -digit groups, gives one a natural distance metric.

## REFERENCES

- Acemoglu, D.; Aghion, P.; Lelarge, C.; Van Reenen, J. and Zilibotti, F., 2007, 'Technology, Information, and the Decentralization of the Firm,' *Quarterly Journal of Economics*, 122 (4), pp. 1759–1799.
- Akerberg, D.; Benkard, C. L.; Berry, S. and Pakes, A., 2007, 'Econometric Tools for Analyzing Market Outcomes,' volume 6 of *Handbook of Econometrics* (Elsevier).
- Acs, Z.; Audretsch, D. and Feldman, M., 1992, 'Real Effects of Academic Research: Comment,' *American Economic Review*, 82 (1), pp. 363–367.
- Adams, J., 1990, 'Fundamental Stocks of Knowledge and Productivity Growth,' *Journal of Political Economy*, 98 (4), pp. 673–702.
- Adams, J. and Clemmons, J., 2008, 'Science and Industry: Tracing the Flow of Basic Research through Manufacturing and Trade,' *Economics of Innovation and New Technology*, 17 (5), pp. 473–495.
- Adams, J. and Jaffe, A., 1996, 'Bounding the Effects of R&D: An Investigation Using Matched Establishment–Firm Data,' *RAND Journal of Economics*, 27 (4), pp. 700–721.
- Agrawal, A.; Cockburn, I. and McHale, J., 2006, 'Gone but not Forgotten: Knowledge Flows, Labor Mobility, and Enduring Social Relationships,' *Journal of Economic Geography*, 6 (5), pp. 571–591.
- Aldieri, L. and Cincera, M., 2009, 'Geographic and Technological R&D Spillovers within the Triad: Micro Evidence from US Patents,' *Journal of Technology Transfer*, 34 (2), pp. 196–211.
- Almeida, P. and Kogut, B., 1999, 'Localization of Knowledge and the Mobility of Engineers in Regional Networks,' *Management Science*, 45 (7), pp. 905–917.
- Arellano, M. and Bond, S., 1991, 'Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,' *Review of Economic Studies*, 58 (2), pp. 277–297.
- Arellano, M. and Bover, O., 1995, 'Another Look at the Instrumental Variable Estimation of Error–Components Models,' *Journal of Econometrics*, 68 (1), pp. 29–51.
- Bartelsman, E.; Becker, R. and Gray, W., 2000, 'NBER–CES Manufacturing Industry Database,' NBER.

- Belenzon, S. and Schankerman, M., 2013, 'Spreading the Word: Geography, Policy, and Knowledge Spillovers,' *Review of Economics and Statistics*, 95 (3), pp. 884–903.
- Bernstein, J., 1988, 'Costs of Production, Intra–and Interindustry R&D Spillovers: Canadian Evidence,' *Canadian Journal of Economics*, 21 (2), pp. 324–347.
- Bernstein, J. and Nadiri, M., 1989, 'Research and Development and Intra–Industry Spillovers: An Empirical Application of Dynamic Duality,' *Review of Economic Studies*, 56 (2), pp. 249–267.
- Berry, S., 1994, 'Estimating Discrete–Choice Models of Product Differentiation,' *RAND Journal of Economics*, 25 (2), pp. 242–262.
- Berry, S.; Levinsohn, J. and Pakes, A., 1995, 'Automobile Prices in Market Equilibrium,' *Econometrica*, 63 (4), pp. 841–890.
- Bloom, N.; Schankerman, M. and Van Reenen, J., 2013, 'Identifying Technology Spillovers and Product Market Rivalry,' *Econometrica*, 81 (4), pp. 1347–1393.
- Blundell, R. and Bond, S., 1998, 'Initial Conditions and Moment Restrictions in Dynamic Panel Data Models,' *Journal of Econometrics*, 87 (1), pp. 115–143.
- Blundell, R. and Bond, S., 2000, 'GMM Estimation with Persistent Panel Data: an Application to Production Functions,' *Econometric Reviews*, 19 (3), pp. 321–340.
- Bresnahan, T.; Brynjolfsson, E. and Hitt, L., 2002, 'Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm–Level Evidence,' *Quarterly Journal of Economics*, 117 (1), pp. 339–376.
- Buettner, T., 2004, 'R&D and the Dynamics of Productivity,' mimeo, London School of Economics.
- Cairncross, F., 2001, *The Death of Distance: How the Communications Revolution is Changing our Lives* (Harvard Business Press, Cambridge).
- Caves, D.; Christensen, L. and Diewert, W., 1982, 'Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers,' *Economic Journal*, 92 (365), pp. 73–86.
- Chang, Y. and Hong, J., 2006, 'Do Technological Improvements in the Manufacturing Sector Raise or Lower Employment?' *American Economic Review*, 96 (1), pp. 352–368.
- Cincera, M., 2005, 'Firm's Productivity Growth and R&D Spillovers: An Analysis of Alternative Technological Proximity Measures,' *Economics of Innovation and New Technology*, 14 (8), pp. 657–682.
- Combes, P. and Duranton, G., 2006, 'Labour Pooling, Labour Poaching, and Spatial Clustering,' *Regional Science and Urban Economics*, 36 (1), pp. 1–28.

- Cragg, J. G. and Donald, S. G., 1993, 'Testing Identifiability and Specification in Instrumental Variable Models,' *Econometric Theory*, 9 (02), pp. 222–240.
- De Loecker, J., 2011, 'Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity,' *Econometrica*, 79 (5), pp. 1407–1451.
- Eaton, J. and Kortum, S., 1996, 'Trade in Ideas Patenting and Productivity in the OECD,' *Journal of International Economics*, 40 (3), pp. 251–278.
- Fallick, B.; Fleischman, C. A. and Rebitzer, J. B., 2006, 'Job-Hopping in Silicon Valley: Some Evidence Concerning the Microfoundations of a High-Technology Cluster,' *Review of Economics and Statistics*, 88 (3), pp. 472–481.
- Freeman, L. M., 2008, 'Job Hopping, Earnings Dynamics, and Industrial Agglomeration in the Software Publishing Industry,' *Journal of Urban Economics*, 64 (3), pp. 590–600.
- Gandhi, A.; Navarro, S. and Rivers, D., 2013, 'On the Identification of Production Functions: How Heterogeneous is Productivity?' mimeo.
- Gasmi, F.; Laffont, J. and Vuong, Q., 2005, 'Econometric Analysis of Collusive Behavior in a Soft-Drink Market,' *Journal of Economics and Management Strategy*, 1 (2), pp. 277–311.
- Goto, A. and Suzuki, K., 1989, 'R&D Capital, Rate of Return on R&D Investment and Spillover of R&D in Japanese Manufacturing Industries,' *Review of Economics and Statistics*, 71 (4), pp. 555–564.
- Greenstone, M.; Hornbeck, R. and Moretti, E., 2010, 'Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings,' *Journal of Political Economy*, 118 (3), pp. 536–598.
- Griffith, R.; Harrison, R. and Van Reenen, J., 2006, 'How Special Is the Special Relationship? Using the Impact of US R&D Spillovers on UK Firms as a Test of Technology Sourcing,' *American Economic Review*, 96 (5), pp. 1859–1875.
- Griffith, R.; Lee, S. and Van Reenen, J., 2011, 'Is Distance Dying at Last? Falling Home Bias in Fixed Effects Models of Patent Citations,' *Quantitative Economics*, 2 (2), pp. 211–249.
- Griliches, Z., 1979, 'Issues in Assessing the Contribution of Research and Development to Productivity Growth,' *Bell Journal of Economics*, 10 (1), pp. 92–116.
- Griliches, Z., 1992, 'The Search for R&D Spillovers,' *Scandinavian Journal of Economics*, 94, pp. 29–47.
- Griliches, Z. and Lichtenberg, F., 1984, 'Interindustry Technology Flows and Productivity Growth: A Reexamination,' *Review of Economics and Statistics*, 66 (2), pp. 324–329.

- Hall, B., 1990, 'The Manufacturing Sector Master File: 1959-1987,' NBER Working Paper No. 3366.
- Hall, B.; Jaffe, A. and Trajtenberg, M., 2001, 'The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools,' NBER Working Paper No. 8498.
- Hall, B.; Jaffe, A. and Trajtenberg, M., 2005, 'Market Value and Patent Citations,' *RAND Journal of Economics*, 36 (1), pp. 16–38.
- Jaffe, A., 1986, 'Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value,' *American Economic Review*, 76 (5), pp. 984–1001.
- Jaffe, A., 1989, 'Real Effects of Academic Research,' *American Economic Review*, 79 (5), pp. 957–970.
- Jaffe, A. and Trajtenberg, M., 2005, *Patents, Citations, and Innovations: A Window on the Knowledge Economy* (MIT press, Cambridge).
- Jaffe, A.; Trajtenberg, M. and Henderson, R., 1993, 'Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations,' *Quarterly Journal of Economics*, 108 (3), pp. 577–598.
- Kadiyali, V., 1996, 'Entry, its Deterrence, and its Accommodation: A Study of the US Photographic Film Industry,' *RAND Journal of Economics*, 27 (3), pp. 452–478.
- Keller, W., 2002, 'Geographic Localization of International Technology Diffusion,' *American Economic Review*, 92 (1), pp. 120–142.
- Klette, T., 1996, 'R&D, Scope Economies, and Plant Performance,' *RAND Journal of Economics*, 27 (3), pp. 502–522.
- Levin, R. and Reiss, P., 1988, 'Cost-Reducing and Demand-Creating R&D with Spillovers,' *RAND Journal of Economics*, 19 (4), pp. 538–556.
- Levinsohn, J. and Petrin, A., 2003, 'Estimating Production Functions Using Inputs to Control for Unobservables,' *Review of Economic Studies*, 70 (2), pp. 317–341.
- Lychagin, S.; Pinkse, J.; Slade, M. and Van Reenen, J., 2010, 'Spillover in Space: Does Geography Matter?' NBER Working Paper No. 16188.
- Marschak, J. and Andrews Jr, W., 1944, 'Random Simultaneous Equations and the Theory of Production,' *Econometrica*, 13 (1), pp. 143–205.
- Olley, G. and Pakes, A., 1996, 'The Dynamics of Productivity in the Telecommunications Equipment Industry,' *Econometrica*, 64 (6), pp. 1263–1297.
- Orlando, M., 2004, 'Measuring Spillovers from Industrial R&D: On the Importance of Geographic and Technological Proximity,' *RAND Journal of Economics*, 35 (4), pp. 777–786.

- Ornaghi, C., 2006, 'Spillovers in Product and Process Innovation: Evidence from Manufacturing Firms,' *International Journal of Industrial Organization*, 24 (2), pp. 349–380.
- Pinkse, J. and Slade, M., 2004, 'Mergers, Brand Competition, and the Price of a Pint,' *European Economic Review*, 48 (3), pp. 617–643.
- Pinkse, J.; Slade, M. and Brett, C., 2002, 'Spatial Price Competition: A Semiparametric Approach,' *Econometrica*, 70 (3), pp. 1111–1153.
- Roodman, D., 2006, 'How to Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata.' *Technical report*, Center for Global Development.
- Rosenthal, S. and Strange, W., 2003, 'Geography, Industrial Organization, and Agglomeration,' *Review of Economics and Statistics*, 85 (2), pp. 377–393.
- Scherer, F., 1982, 'Inter-Industry Technology Flows and Productivity Growth,' *Review of Economics and Statistics*, 64 (4), pp. 627–634.
- Slade, M., 1995, 'Product Rivalry with Multiple Strategic Weapons: An Analysis of Price and Advertising Competition,' *Journal of Economics and Management Strategy*, 4 (3), pp. 445–476.
- Spence, M., 1984, 'Cost Reduction, Competition, and Industry Performance,' *Econometrica*, 52 (1), pp. 101–21.
- Staiger, D. and Stock, J. H., 1997, 'Instrumental Variables Regression with Weak Instruments,' *Econometrica*, 65 (3), pp. 557–586.
- Stock, J. H. and Yogo, M., 2005, 'Testing for Weak Instruments in Linear IV Regression,' *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 1.
- Thompson, P., 2006, 'Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-Added Citations,' *Review of Economics and Statistics*, 88 (2), pp. 383–388.
- Villalonga, B., 2004, 'Diversification Discount or Premium? New Evidence from the Business Information Tracking Series,' *Journal of Finance*, 59 (2), pp. 479–506.
- Wilson, D., 2009, 'Beggar Thy Neighbor? The In-State, Out-of-State, and Aggregate Effects of R&D Tax Credits,' *Review of Economics and Statistics*, 91 (2), pp. 431–436.

## DATA APPENDIX

*The patents and Compustat databases*

We obtained our dataset by merging firm-level accounting data from S&P Compustat with the NBER Patents database. We started with the universe of Compustat manufacturing firms with headquarters based in the U.S.<sup>42</sup> We kept firms that had non-zero sales in at least one year between 1980 and 2000 and appeared in the Compustat Segment (“line of business”) Database. The Segment Database allocates firm sales into four digit industries each year using firm’s descriptions of their sales by lines of business. See [Villalonga \(2004\)](#) for a more detailed description. This left us with 6,890 firms in our sample.

Next, we matched the remaining firms into the NBER data on patents. The NBER patents database provides detailed patenting and citation information for around 2,500 firms (as described in [Hall et al., 2005](#); [Jaffe and Trajtenberg, 2005](#)). We used patent data between 1970 to 2000<sup>43</sup>, and kept only patenting firms. The accounting and the patent data do not have to be concurrent. For example, a firm which patented in 1985, 1988 and 1989, had Segment data from 1993 to 1997, and accounting data from 1980 to 1997 would be kept in our dataset for the period 1980 to 1997. After this stage, our sample contains data on 1,542 firms.

The dataset was further cleaned to remove accounting years with extremely large jumps in sales and gaps in firms’ annual reports signaling merger and acquisition activity. We also used the Compustat footnotes that indicate significant mergers. When we remove a year we treat the firm as a new entity and give it a new identifier (and therefore a new fixed effect) even if the firm identifier (GVKEY reference) in Compustat remained the same. This is more general than including a full set of firm fixed effects as we are allowing the fixed effect to change over time.

This left a final sample of 1,383 firms to estimate the model on with accounting data for at least some of the period 1980 to 2000 and patenting data for at least some of the period between 1970 and 2000. The panel is unbalanced as we keep new entrants and exitors in the sample.

*Variables*

We use deflated sales as our output measure (Compustat mnemonic SALE). The book value of capital is the net stock of property, plant and equipment (PPENT). Employment is the number of employees (EMP). R&D expenditure (XRD) is used to create R&D capital stocks calculated using a perpetual inventory method with a 15% depreciation rate. Material expenses are defined as in [Bresnahan et al. \(2002\)](#): sales less operating income before depreciation (OIBD) less labor expenses (XLR). Labor expenses are not always reported in the Compustat dataset; in such cases we construct labor expenses by taking a product of a firm-specific wage and the number of

<sup>42</sup>We dropped firms with headquarters outside of the U.S. as many of those firms do not have domestic plants and thus face different factor prices. However, including those firms makes little difference to our results.

<sup>43</sup>We dropped pre-1970 data as being too outdated for our 1980s and 1990s accounts data.



employees. The construction of the spillover variables is described in Section IV above in detail. About 95% of the variance of the spillover measures, SpillSIC, SpillTech and SpillGeog is between firm and 5% is within firm. When we include fixed effects we are, of course, relying on the time series variation for identification.

In order to construct the Tornqvist productivity index (15), we use firm and time specific cost shares of inputs. This allows for possible heterogeneity in the production technology across firms and years. We assume that the technology exhibits constant returns to scale, so that the total cost equals total revenue and the shares of labor and intermediate materials equal the respective expenses divided by sales. We do not directly observe capital expenses; the share of capital in total cost is found by subtracting the shares of labor and materials from unity.

### *Geographic coordinates*

We obtain the geographic coordinates of inventors by merging the patents data and the extract from the Geographic Names Information System (GNIS) provided by the U.S. Geological Survey. All inventor records contain address information, in most cases the names of the home state and the city. The GNIS records provide coordinates for all established geographic names.

Most of the time city and state names are not sufficient to uniquely identify a geographic place: a typical city name has a number of duplicates within its state. If these duplicates are located in the same county or within 0.1 degrees of latitude and longitude, we treat them as one entity.

If the duplicated names are located far apart, we give preference to incorporated places. Names of incorporated places are usually more permanent and stable, they are more likely to be used by the U.S. Postal Service. In turn, inventors are more likely to indicate a USPS-preferred city name when they are asked to provide their home address.

Following these rules, we are able to match 98.25% of inventor records that are related to the firms in our sample. To reduce computational burden, we aggregate all geographic data to the county level. For example, we assign identical coordinates to inventors from Long Beach CA and Los Angeles CA: both cities are located in Los Angeles County.

The geographic coordinates of firm headquarters are obtained from the Compustat dataset. The drawback of using the Compustat data is that we observe only the last known locations at the time when the data had been downloaded. Some companies in our sample (e.g. Boeing) relocate their headquarters after 2000; by using headquarter locations from Compustat we misplace these firms in geographic space.

### Deflators

Sales are deflated by the geometric mean of prices of all 4-digit manufacturing products sold by a firm weighted by the product shares in the total sales of the firm:

$$(26) \quad P_{it}^Q = \exp \left( \sum_{\text{SIC}} \frac{\text{SALE}_{\text{SIC},it}}{\text{SALE}_{it}} \ln P_{\text{SIC},t}^Q \right)$$

The total firm sales are broken into 4-digit SIC categories using Compustat segment data. Year-specific product level prices,  $P_{\text{SIC},t}^Q$ , are obtained from the NBER-CES Manufacturing Industry Database ([Bartelsman et al., 2000](#)).

*Capital stock deflator.* To deflate the book value of capital we follow [Hall \(1990\)](#). First, we use the depreciation in year  $t$  to calculate  $A_{it}$ , an average age of firm  $i$ 's working capital stock. Second, we divide the net book value of firm  $i$ 's property, plant and equipment in year  $t$  by the firm-specific capital expenditure deflator that matches the average age of working capital,  $P_{t-A_{it}}^K$ .

Capital expenditure deflator is the geometric mean of investment prices in all 4-digit SIC industries where the firm was active at any year weighted by the share of investment  $I_{\text{SIC},i} / I_i$  made by the firm in each product line in the firm's total investment expenses:

$$(27) \quad P_{it}^K = \exp \left( \sum_{\text{SIC}} \frac{I_{\text{SIC},i}}{I_i} \ln P_{\text{SIC},t}^K \right), \quad \text{where } I_{\text{SIC},i} = \sum_t \frac{I_{\text{SIC},it}}{P_{\text{SIC},t}^K}, \quad I_i = \sum_{\text{SIC}} I_{\text{SIC},i}$$

The industry-specific price of investment  $P_{\text{SIC},t}^K$  is obtained from the NBER-CES database; firm-level capital expenditures are drawn from the Compustat Segments file.

*Materials price deflator.* Material expenses are deflated using industry level material prices from the NBER-CES database. We construct the firm-specific price index for materials by taking the weighted geometric mean of the industry prices:

$$(28) \quad P_{it}^M = \exp \left( \sum_{\text{SIC}} \alpha_{\text{SIC},it} \ln P_{\text{SIC},t}^M \right), \quad \text{where } \sum_{\text{SIC}} \alpha_{\text{SIC},it} = 1$$

The weights  $\alpha_{\text{SIC},it}$  are proportional to sales of the firm's respective products and shares of industry-level material expenses (MAT) in total industry shipments (SHIP):

$$(29) \quad \alpha_{\text{SIC},it} \propto \text{SALE}_{\text{SIC},it} \frac{P_{\text{SIC},t}^M \text{MAT}_{\text{SIC},t}}{\text{SHIP}_{\text{SIC},t}}$$

The weights  $\alpha_{\text{SIC},t}$  are meant to approximate the relative material requirement of each product segment of the firm.

*Wages.* Firm level wage,  $w_{it}$ , is equal to total adjusted<sup>44</sup> labor expense divided by the number of employees:

$$(30) \quad w_{it} = \sum_{\text{SIC}} \left( \frac{w_{\text{SIC},it} L_{\text{SIC},it}}{\gamma_{\text{SIC},t}} \right) / L_{it},$$

We observe neither segment specific labor expenses, nor segment specific employment. Therefore, we need extra assumptions about the structure of firm labor expenses in order to compute wages. We assume that the share of labor expenses in each segment's sales is a product of two components: (1) an industry-level share, found from the NBER-CES data, (2) an unknown firm-specific multiplier:

$$(31) \quad \frac{w_{\text{SIC},it} L_{\text{SIC},it}}{\text{SALE}_{\text{SIC},it}} = \frac{w_{\text{SIC},t} \text{EMP}_{\text{SIC},t}}{\text{SHIP}_{\text{SIC},t}} b_{it}.$$

We also assume that all firms pay the same wage within an industry, i.e.  $w_{\text{SIC},it} = w_{\text{SIC},t}$ . Then, by solving for  $L_{\text{SIC},it}$  and taking a sum across SIC codes, one should obtain the total number of employees,  $L_{it}$ , as a function of  $b_{it}$ . Since  $L_{it}$  is available in the Compustat data, one can compute  $b_{it}$ :

$$(32) \quad b_{it} = L_{it} \left( \sum_{\text{SIC}} \frac{\text{EMP}_{\text{SIC},t}}{\text{SHIP}_{\text{SIC},t}} \text{SALE}_{\text{SIC},it} \right)^{-1}$$

All variables on the right hand side are available in the data; by substituting  $b_{it}$  we find segment-specific labor expenses. Finally, we use these expenses to calculate the firm-specific wage  $w_{it}$ .

*R&D deflator.* R&D expenses are deflated using aggregate input price index for R&D performed by private businesses, published by BEA as a part of the R&D Satellite Account.

---

<sup>44</sup>The adjustment is necessary to account for fringe benefits, which are not included in the NBER-CES payroll data. The adjustment factor  $\gamma_{\text{SIC},t}$  is found from NIPA 2-digit industry accounts, and is equal to the ratio of industry payroll to total industry compensation (see [Chang and Hong, 2006](#), for more details).

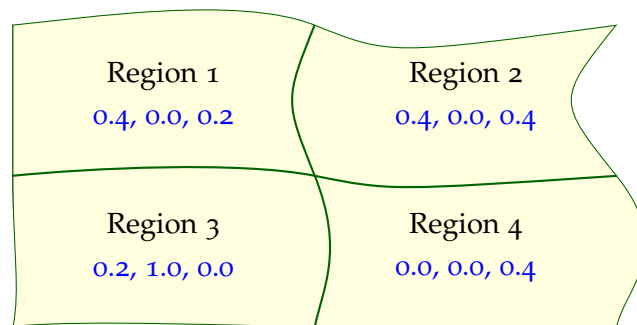


Figure 1  
Construction of Geographic Spillover Measures

There are three firms and four regions. The fraction of inventors belonging to each firm in each region is indicated in the graph, so firm 1 has 40% of its inventors in region 1, 40% in region 2 and 20% in region 3, firm 2 has all inventors in region 3 and firm 3 has 20% in region 1, 40% in region 2 and 40% in region 4.

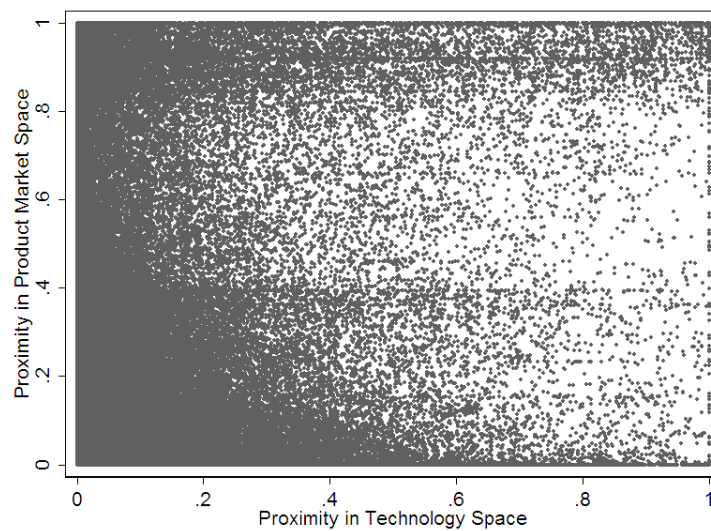


Figure 2  
SIC and TECH Correlations

Each dot represents a  $(w^P, w^T)$  combination for a pair of firms, measuring product and technology proximity respectively.

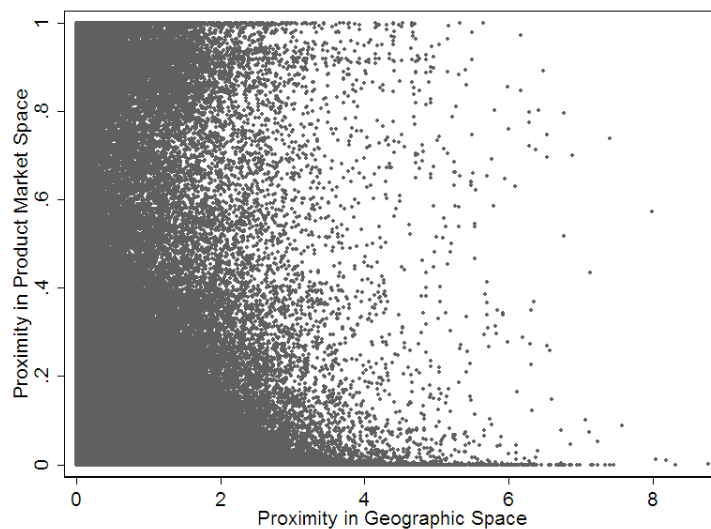


Figure 3  
SIC and GEOG Correlations

Each dot represents a  $(w^P, w^G)$  combination for a pair of firms, measuring product and geographic proximity respectively.

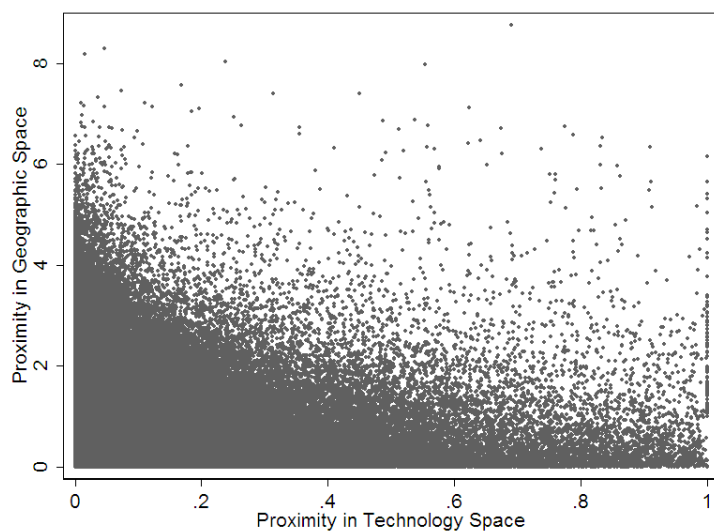


Figure 4  
GEOG and TECH Correlations

Each dot represents a  $(w^G, w^T)$  combination for a pair of firms, measuring geographic and technology proximity respectively.

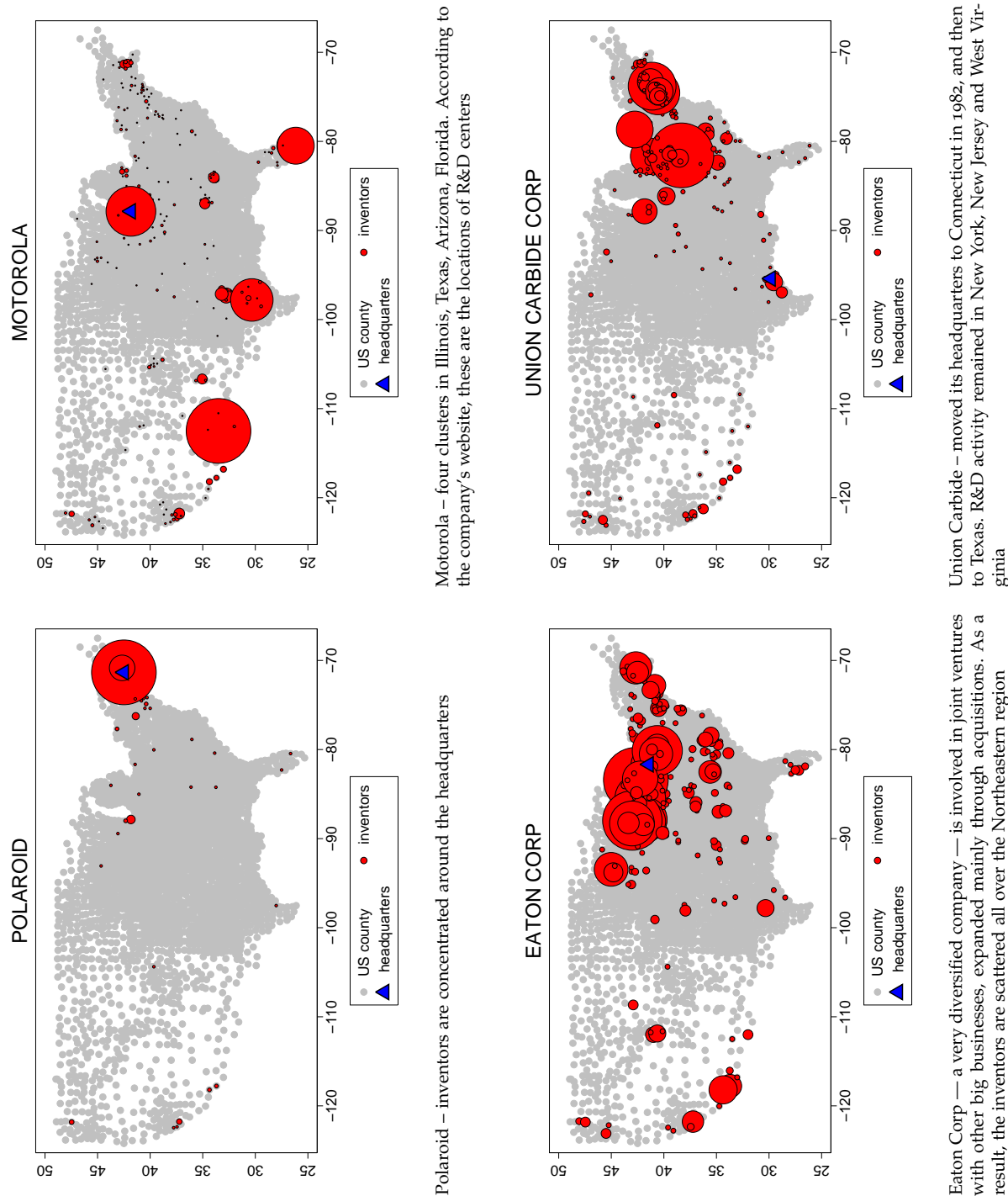


Figure 5  
Inventor Distributions, Selected Firms

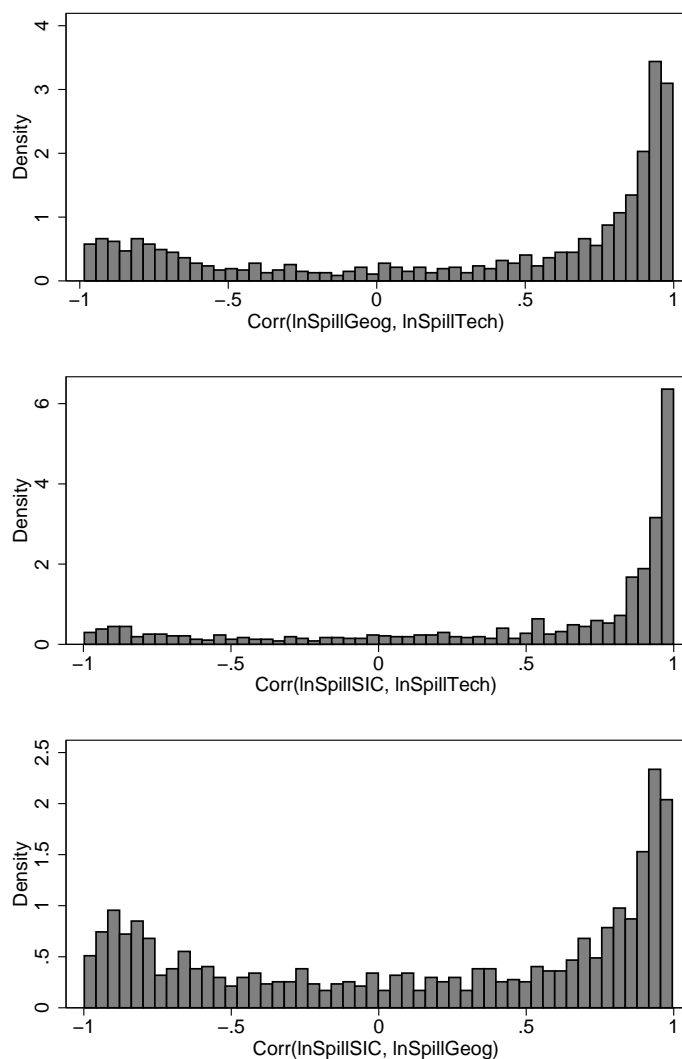


Figure 6  
Within Firm Correlations of Spillover Measures

The graphs display the distributions of within-firm correlations between pairs of spillover measures. Correlations  $\text{corr}(\ln\text{SpillTech}, \ln\text{SpillGeog})$  are computed by removing the respective year means from the technological and the geographic spillover variables and finding their correlation coefficient for every firm with at least five observations in the data.  $\text{corr}(\ln\text{SpillSIC}, \ln\text{SpillGeog})$  and  $\text{corr}(\ln\text{SpillTech}, \ln\text{SpillSIC})$  are found in the analogous way.

TABLE I  
DESCRIPTIVE STATISTICS

Variable	(1) Mean	(2) Median	(3) Std.dev.
Output, $Q$	2,010	279	10,279
Stock of physical capital, $K$	725	74	2,830
Employment, $L$	11,562	2,545	35,754
Materials, $M$	1,128	154	4,688
R&D stock, $S$	254	15	1,317
Technological spillovers, SpillTech	15,094	12,264	11,636
lnSpillTech	9.266	9.414	0.986
Product market spillovers, SpillSIC	15,149	9,574	16,297
lnSpillSIC	8.867	9.167	1.536
Geographic spillovers, SpillGeog	131,747	116,994	87,850
lnSpillGeog	11.525	11.670	0.831

The means, medians and standard deviations are taken over all non-missing observations between 1980 and 2000; values measured in 1987 prices in \$million. The geographic spillover measure is based on inventor locations. The half-life distance of geographic spillovers is set to 200 km.

TABLE II  
CORRELATIONS BETWEEN THE R&D STOCK MEASURES

Correlation between	(1) SIC–Tech	(2) SIC–Geog	(3) Geog–Tech	(4) Own–SIC	(5) Own–Tech	(6) Own–Geog
X	0.475	0.144	0.332	0.212	0.342	0.190
$\Delta X$	0.530	0.224	0.299	0.280	0.355	0.197
Covariance						
X	0.720	0.183	0.272	0.726	0.753	0.353

All correlations are significant at the 1% level. X denotes lnSpillSIC, lnSpillTech, lnSpillGeog, or lnOwnS.  $\Delta X$  is defined as the residual from a regression with firm fixed effects and time dummies.

TABLE III  
DISTANCE BETWEEN INVENTORS AND HEADQUARTERS

Percentile	0%	20%	40%	60%	80%	100%
Distance	0 km	0 km	52 km	414 km	1,603 km	8,097 km

Each quintile shows the aggregate percentage of inventors that are located at various distances from their headquarters. Approximately 99.94% of inventors are located within a 4,300 km radius from their headquarters (roughly, the distance between Boston and San Francisco); the remaining 0.06% involve labs or headquarters in Alaska or Hawaii.



TABLE IV  
OLS INDEX NUMBER ESTIMATES, SINGLE PROXIMITY MEASURES

Dependent variable is TFP										
	(1)		(2)		(3)		(4)		(5)	
	None		SpillSIC		SpillTech		SpillGeog		All Three	
lnOwnS	0.050	(5.45)	0.028	(3.64)	0.011	(1.61)	0.032	(4.12)	0.001	(0.10)
lnSpillSIC			0.340	(6.67)					0.123	(3.34)
lnSpillTech					0.921	(7.45)			0.654	(6.47)
lnSpillGeog							1.113	(5.10)	0.835	(4.36)
# obs.	17,010		17,010		17,010		17,010		17,010	

This table reports OLS estimates of TFP on the demeaned explanatory variables. Year fixed effects are included. t-statistics are in parentheses (using clustering by firm). The geographic spillover variable is based on inventor locations and the exponential-decay distance function; the spillover is set to decay at a rate of 50% per 200km. lnOwnS = 0 if OwnS = 0. Otherwise, lnOwnS = ln(OwnS). A dummy variable that indicates that OwnS = 0 is included.

TABLE V  
OLS INDEX NUMBERS, INVENTORS AND HEADQUARTERS, DIFFERENT C FUNCTIONS

Dependent variable is TFP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Based on locations of:	Inventors			Headquarters		Both	
Distance Fn (C)	Exp	Corr	Both	Exp	Corr	Exp	Corr
lnOwnS	0.001 (0.10)	0.005 (0.72)	0.001 (0.12)	0.002 (0.31)	0.006 (0.79)	0.000 (0.06)	0.004 (0.53)
lnSpillSIC	0.123 (3.34)	0.142 (3.71)	0.126 (3.36)	0.122 (3.36)	0.141 (3.31)	0.118 (3.26)	0.146 (3.34)
lnSpillTech	0.654 (6.47)	0.704 (6.53)	0.662 (6.46)	0.689 (6.76)	0.879 (7.06)	0.655 (6.56)	0.802 (6.97)
lnSpillGInvExp	0.835 (4.36)		0.912 (4.46)			0.427 (2.70)	
lnSpillGInvCorr		0.202 (3.25)	-0.058 (-1.29)				0.201 (2.84)
lnSpillGHeadExp				0.444 (4.34)		0.295 (3.05)	
lnSpillGHeadCorr					0.028 (1.75)		0.015 (1.00)
# obs.	17,010	16,982	16,982	17,010	14,772	17,010	14,748

This table reports OLS estimates of TFP on the demeaned explanatory variables. Year fixed effects are included. t-statistics are in parentheses (clustered by firm). Exp denotes an exponentially decaying distance function; the spillover is set to decay at a rate of 50% per 200km. Corr denotes a 0/1 distance function (same or different geographic region). lnOwnS = 0 if OwnS = 0. Otherwise, lnOwnS = ln(OwnS). A dummy variable indicating that OwnS = 0 is included. Variables lnSpillGInvExp and lnSpillGInvCorr use inventor locations, while lnSpillGHeadExp and lnSpillGHeadCorr use headquarter locations.

TABLE VI  
GMM INDEX NUMBER ESTIMATES

Dependent variable is TFPL			
	Arellano & Bond	Dynamic GMM	COMFAC
	(1)	(2)	(3)
TFPL <sub>t-1</sub>		0.406 (4.07)	0.390 (4.03)
lnOwnS	0.076 (1.37)	0.003 (0.06)	0.006 (0.12)
lnOwnS <sub>t-1</sub>		0.068 (1.51)	
lnSpillSIC	0.082 (1.92)	-0.107 (-0.52)	0.085 (1.83)
lnSpillSIC <sub>t-1</sub>		0.266 (0.78)	
lnSpillTech	0.196 (1.96)	1.622 (2.24)	0.322 (2.86)
lnSpillTech <sub>t-1</sub>		-1.875 (-2.13)	
lnSpillGeog	0.420 (2.82)	-2.549 (-1.12)	0.381 (2.03)
lnSpillGeog <sub>t-1</sub>		3.025 (1.33)	
# instruments	129	128	128
Comfac Rest p-val			0.03
Hansen test	$\chi^2(104) = 131$	$\chi^2(103) = 132$	
Hansen p-val	0.04	0.06	
AR(1) p-val	0.00	0.00	
AR(2) p-val	0.10	0.02	
# obs.	15,039	13,272	13,272

TFPL denotes TFP in levels as defined in equation (18). Year fixed effects are included; t-statistics are reported in parentheses (clustered by firm). The geographic spillover variable is based on inventor locations and the exponential-decay distance function; the spillover is set to decay at a rate of 50% per 200km. lnOwnS = 0 if OwnS = 0. Otherwise, lnOwnS = ln(OwnS). A dummy variable indicating that OwnS = 0 is included; its lag is included in column 2. In all three specifications we use moment equations in first differences. Lags 2–4 of lnOwnS and the zero R&D dummy are used as GMM-style instruments for the differences of these variables and the lagged difference of TFP. The instruments are defined as in Roodman (2006) by creating a separate set of lags for each time period and replacing missing values with zeros. Differences of the spillover variables are treated as exogenous.

TABLE VII  
INTERACTIONS

Dependent variable is TFP								
	(1)		(2)		(3)		(4)	
lnOwnS	0.001	(0.10)	0.001	(0.10)	0.001	(0.08)	0.001	(0.12)
lnSpillSIC	0.123	(3.34)	0.130	(1.06)	0.122	(3.31)	0.150	(1.07)
lnSpillTech	0.654	(6.47)	0.654	(6.47)	0.635	(3.83)	0.619	(3.39)
lnSpillGeog	0.835	(4.36)	0.837	(4.42)	0.825	(4.52)	0.830	(4.49)
SICGeog			-0.001	(-0.05)			-0.003	(-0.21)
TechGeog					0.003	(0.15)	0.005	(0.24)
# obs.	17,010		17,010		17,010		17,010	

This table reports OLS estimates of TFP on the demeaned explanatory variables. Year fixed effects are included; t-statistics are reported in parentheses (clustered by firm). The geographic spillover variable is based on inventor locations and the exponential-decay distance function; the spillover is set to decay at a rate of 50% per 200km. lnOwnS = 0 if OwnS = 0. Otherwise, lnOwnS = ln(OwnS). A dummy variable indicating that OwnS = 0 is included. SICGeog denotes the interaction of lnSpillSIC and lnSpillGeog. TechGeog denotes the interaction of lnSpillTech and lnSpillGeog.

TABLE VIII  
HALF-LIFE ESTIMATES

Dependent variable is TFP		
	(1)	(2)
Half-life, km	259 (2.48)	
Half-life (1980–1990), km		247 (2.63)
Half-life (1991–2000), km		254 (2.63)
lnOwnS	0.001 (0.11)	0.001 (0.11)
lnSpillSIC	0.124 (3.37)	0.128 (3.49)
lnSpillTech	0.658 (6.52)	0.659 (6.47)
lnSpillGeog	0.916 (3.99)	0.923 (4.38)
# obs.	17,010	17,010

This table reports the estimates of spillover half-life. Year fixed effects are included; t-statistics are reported in parentheses (clustered by firm). The geographic spillover variable is based on inventor locations and the exponential-decay distance function with estimated half-life. lnOwnS = 0, if OwnS = 0. Otherwise, lnOwnS = ln(OwnS). A dummy variable indicating that OwnS = 0 is included.

TABLE IX  
OLS FOUR-FACTOR TFP ESTIMATES, SINGLE PROXIMITY MEASURES

Dependent variable is four-factor TFP								
	(1)		(2)		(3)		(4)	
	SpillSIC		SpillTech		SpillGeog		All Three	
lnSpillSIC	0.350	(6.57)					0.116	(3.19)
lnSpillTech			0.886	(7.24)			0.601	(6.17)
lnSpillGeog					1.127	(5.13)	0.800	(4.20)
# obs.	17001		17001		17001		17001	

This table reports OLS estimates of four-factor TFP on the demeaned explanatory variables. The four factors are labor, capital, intermediate inputs and R&D stock. The share of R&D is calculated as a ratio of R&D expenditures to sales. It is assumed that 50% of R&D expenses account for labor, and another 50% account for intermediate materials. We adjust the respective expenditures to avoid double accounting. Year fixed effects are included; t-statistics are reported in parentheses (clustered by firm). The geographic spillover variable is based on inventor locations and the exponential-decay distance function; the spillover is set to decay at a rate of 50% per 200km.

TABLE X  
OLS INDEX NUMBER ESTIMATES USING DIFFERENT MEASURES OF SIZE

Dependent variable is TFP			
	(1)	(2) lnSpillK	(3) lnSpillL
lnOwnS	0.001 (0.10)	0.000 (0.02)	0.000 (0.06)
lnSpillSIC	0.123 (3.34)	0.118 (3.18)	0.123 (3.32)
lnSpillTech	0.654 (6.47)	0.667 (6.60)	0.662 (6.57)
lnSpillGeog	0.835 (4.36)	0.718 (4.08)	0.762 (4.23)
lnSpillK		0.177 (1.80)	
lnSpillL			0.160 (1.54)
# obs.	17010	17010	17010

This table reports OLS estimates of TFP on the demeaned explanatory variables. Year fixed effects are included; t-statistics are reported in parentheses (clustered by firm). The geographic spillover variable is based on inventor locations and the exponential-decay distance function; the spillover is set to decay at a rate of 50% per 200km. lnOwnS = 0, if OwnS = 0. Otherwise, lnOwnS = ln(OwnS). A dummy variable indicating that OwnS = 0 is included. lnSpillK and lnSpillL denote the measures of local manufacturing activity based on capacity and employment, respectively, using geographic R&D weights.

TABLE XI  
OLS PRODUCTION FUNCTIONS, INVENTORS & HEADQUARTERS, DIFFERENT C FUNCTIONS

Dependent variable is lnQ							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Based on locations of:	Inventors			Headquarters		Both	
Distance Fn (C)	Exp	Corr	Both	Exp	Corr	Exp	Corr
lnK	0.052 (3.67)	0.054 (3.76)	0.053 (3.73)	0.052 (3.65)	0.056 (3.56)	0.051 (3.59)	0.052 (3.30)
lnL	0.316 (14.10)	0.310 (13.39)	0.315 (14.08)	0.313 (14.08)	0.296 (12.13)	0.316 (14.25)	0.301 (12.42)
lnM	0.615 (24.97)	0.623 (24.68)	0.615 (25.00)	0.616 (24.84)	0.638 (25.17)	0.613 (24.88)	0.635 (25.08)
lnOwnS	0.005 (0.78)	0.008 (1.24)	0.005 (0.80)	0.006 (1.03)	0.007 (1.04)	0.005 (0.84)	0.006 (0.93)
lnSpillSIC	0.070 (2.01)	0.084 (2.35)	0.071 (2.01)	0.067 (1.99)	0.084 (2.08)	0.065 (1.92)	0.088 (2.14)
lnSpillTech	0.627 (6.40)	0.669 (6.45)	0.639 (6.41)	0.656 (6.68)	0.802 (6.99)	0.630 (6.52)	0.750 (6.80)
lnSpillGInvExp	0.765 (4.44)		0.867 (4.67)			0.352 (2.40)	
lnSpillGInvCorr		0.167 (3.01)	-0.078 (-1.68)				0.152 (2.40)
lnSpillGHeadExp				0.422 (4.61)		0.301 (3.36)	
lnSpillGHeadCorr					0.025 (1.76)		0.016 (1.13)
# obs.	17,010	16,982	16,982	17,010	14,772	17,010	14,748

Firm and year fixed effects included; t-statistics are reported in parentheses (clustered by firm). Exp denotes an exponentially decaying distance function; the respective geographic spillover variables are set to decay at a rate of 50% per 200km. Corr denotes a 0/1 distance function (same or different geographic region). lnOwnS = 0 if OwnS = 0. Otherwise, lnOwnS = ln(OwnS). A dummy variable indicating that OwnS = 0 is included. Variables lnSpillGInvExp and lnSpillGInvCorr use inventor locations, while lnSpillGHeadExp and lnSpillGHeadCorr use headquarter locations.